Author:
Liddle, Emma

Title:
Kinetics and Selectivity of an Isolated Dehydratase Domain from a Fungal Polyketide Synthase

Kinetics and Selectivity of an Isolated Dehydratase Domain from a Fungal Polyketide Synthase

Emma C. Liddle

A thesis submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

University of Bristol

School of Chemistry

Cantock's Close

Bristol, BS8 1TS

March 2018

Abstract

Kinetics and Selectivity of an Isolated Dehydratase domain from a Fungal Polyketide Synthase

Previous work in the field has shown it is possible to isolate catalytic domains from polyketide synthases (PKS) which are able to carry out their programmed catalytic function independently from the complete PKS enzyme. A number of dehydratase (DH) domains have been isolated from modular polyketide synthase (mPKS) systems and their substrate selectivity characterised. Until now, there has been no published study of a DH domain from an iterative polyketide synthase (iPKS). Previous work within the Cox group has shown that it is possible to isolate and characterise the selectivity of the enoyl reductase (ER) domain from squalestatin tetraketide synthase (SQTKS). In order to further the understanding of the SQTKS enzyme, the isolated DH enzyme was produced and tested *in vitro* with substrates designed to test the stereoselectivity of the enzyme.

The assays developed and carried out for this work utilised LCMS to measure the initial rate of the enzyme reaction in order to assess the selectivity of the DH domain. The results demonstrated that the DH domain is highly selective and the correct stereochemistry of the α and β positions is vital for efficient substrate turnover. A second set of assays investigated the potential for non-native substrates to act as inhibitors for the DH. None of the non-substrates inhibited the dehydration reaction, indicating that the substrates were unable to enter the active site.

In the absence of a crystal structure, a model of the SQTKS DH domain was produced using crystal structures of isolated modular DH enzymes. Docking studies *via* a homologous model were undertaken to rationalise the selectivity of the SQTKS DH. Docking of the substrate into the active site of the model showed the programming of the DH domain arises from the position of the two catalytic amino acids, H44 and D235. The position of these amino acids dictates which substrates are tolerated and dehydrated. Through combining the results from the ER and DH assays, it is possible to elucidate the stereo selectivity of almost all of the steps of the SQTKS and show that these are identical to those of mammalian fatty acid synthase (FAS). The significance of these results and their relevance to the field will be discussed.

Acknowledgements

Firstly, I would like to thank Professor Russell Cox for giving me the opportunity to undertake this PhD and for all his guidance over the past four years. I would like to thank Russell and the EPSRC for supplying the funding for my PhD. Secondly, I would also like to thank Professor Chris Willis for offering additional support after Russell's move to Hannover, Germany.

I am forever grateful for my parents unwavering support through my seemingly never ending studies. Thank you for putting up with my waffling and ranting about proteins and synthesis. Without your support I could not have made it where I am today. I promise I will never ask you to read any of my reports ever again.

I am indebted to the numerous members of the Cox group, both past and present for making the lab a much more fun place to be, even when lab work was going wrong. I would like to especially thank Katherine, Kate and Claudio for their support and encouragement and for making me laugh when I needed it the most. I would like to thank Lina for all of our wonderful chats about recipes and for always being there with a warm smile. Outside of the lab, I have made some great friends. To Bryony and Steph, thank you for all your pep talks and copious amounts of red wine. To Tim Harrison, thank you for running ChemLabs and letting me escape my lab and share my passion for science with others.

My thanks go to Dr Alan Scott for teaching me all he knows about protein expression: without your training, this project would not have succeeded. My thanks go to all the members of the Willis group, for welcoming the Cox group into their meetings and for supplying some wonderful cakes during group meetings. I would also like to thank the staff from the analytical services: Dr. Craig Butts, Rose Silvester, Paul Lawrence and Tom Leman from NMR, and Dr. Paul Gates in mass spec.

Last and by no means least, I would like to say a massive thank you to Doug, for doing all the washing and cooking and for putting up with all my ups and downs over the last few months. Thank you for your unwavering support and love, it has kept me sane over the past three years.

Author's Declaration

The work described in this thesis was carried out in the School of Chemistry, University of Bristol under the supervision of Professor C. L. Willis and Professor R. J. Cox between September 2012 and March 2018. The work is original except where indicated by reference and has not been submitted for any other degree. The views expressed are those of the author and in no way represent those of the University of Bristol.

Emma C. Liddle

March 2018

Abbreviations and Units

Å	Angstrom
ACP	Acyl carrier protein
ARO	Aromatase
Asp	Aspartate
AT	Acetyl transferase
Bor	Borrelidin
BSFTA	Bis(trimethylsilyl)trifluoroacetamide
Bu_2BOTf	Dibutylboryl trifluoromethanesulfonate
br	Broad
CHC	Cyclohexanecarboxylate
C-MeT	C-methyltransferase
CoA	Coenzyme A
COSY	Homonuclear correlation spectroscopy
Cur	Curacin
D	Aspartic acid
d.r.	Diastereomeric ratio
DAMP	Deoxyadenosine 5'-monophosphate
DCC	N,N'-Dicyclohexylcarbodiimide
DCM	Dichloromethane
dd	Doublets of doublets
DEBS	Deoxyerythronuclide B synthase
Dfire	All-atom distance-dependent statistical potential
DH	Dehydratase
DHCHC	Dihydroxycyclohexanecarboxylate
DMAP	4-Dimethylaminopyridine
DMBS	Dimethylbassianin Synthase
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDCI	1-Ethyl-3-(3-dimethyiaminopropyl)carbodiimide
ER	Enoyl reductase
Ery	Erythromycin
eq	Equivalents
FA	Fatty acids
FAS	Fatty acid synthase(s)
FPLC	Fast protein liquid chromatography
GC-MS	Gas-chromatograph/Mass spectrometer

GMQE	Global model quality estimation
H or (His)	Histidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMG	hydroxymethyl glutamate
HPLC	High performance liquid chromatography
HR	Highly reducing
Hz	Hertz
iPKS	Iterative polyketide synthase(s)
IPTG	Isopropyl β-D-1-thiogalactopyranoside
J	Coupling constant
KDa	Kilo Dalton
KR	Ketoreductase
KS	Ketosynthase
LB	Luria Bertain media
LCMS	Liquid chromatography/mass spectrometry
LDA	Lithium di-isopropylamide
LDKS	Lovastatin diketide synthase
LNKS	Lovastatin nonaketide synthase
M	Molar
m	Multiplet
m/z	Mass to charge ratio
MAT	Malonyl-acetyl transferase
MCAT	Malonyl-CoA:ACP
mFAS	Mammalian fatty acid synthase(s)
mg	Milligram
mL	Milli-litre
mM	Milli-molar
mp	Melting point
mPKS	Modular polyketide synthase(s)
MPT	Malonyl-palmitoyl transacylase
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
NR-iPKS	Non-reducing iterative polyketides
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthase
OD	Optical density
p	Pentet
PDB	Protein data bank

PicTE	Picromycin thiolesterase
Pik	Picromycin
PK	Polyketide
PKS	Polyketide synthase(s)
PT	Product template
q	Quartet
Rif	Rifamycin
RMSD	Root mean squared deviation
RT	Room temperature
s	Singlet
SAM	*S*-adenosyl methionine
SAT	Starter-unit acyl transferase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIM	Selected ion monitoring
SNAC	*N*-acetylcysteamine
SQTKS	Squalestatin tetraketide synthases
TBDMS	*tert*-Butyldimethylsilyl
TE	Thiolesterase
TEMED	Tetramethylethylenediamine
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxidanyl
TENS	Tenellin synthase
TIC	Total ion current
TLC	Thin layer chromatography
Tyl	Tylosin
UV	Ultra violet
ΨCMeT	Pseudo-methyl transferase
ΨKR	Pseudo-ketoreductase

Table of Contents

Chapter 1 - Introduction

1.1 Natural products

Secondary metabolite natural products are a category of compounds which are of great interest due to the significant role they play in agrochemical research and drug development. The biosynthetic origin dictates the classification of secondary metabolites. The main classifications are terpenes, polyketides, peptides and alkaloids (Figure 1).[1] Terpenes are derived from 5-carbon isoprene units,[1] of which paclitaxel (taxol) **1** is a well-known example as it can be used to treat a range of cancers such as breast and lung cancer.[2] Peptides are another class, which are made of amino acids and can be produced ribosomally or by a dedicated multifunctional enzyme called a non-ribosomal peptide synthetase (NRPS).[1] Vancomycin **2** is an example of a non-ribosomal peptide (NRP), which has antibiotic properties.[3] Alkaloids as a class of natural products vary greatly in structure, but one commonality is the presence of a basic nitrogen atom.[1] One well known alkaloid is morphine **3**, which possesses pain relieving properties.[4]

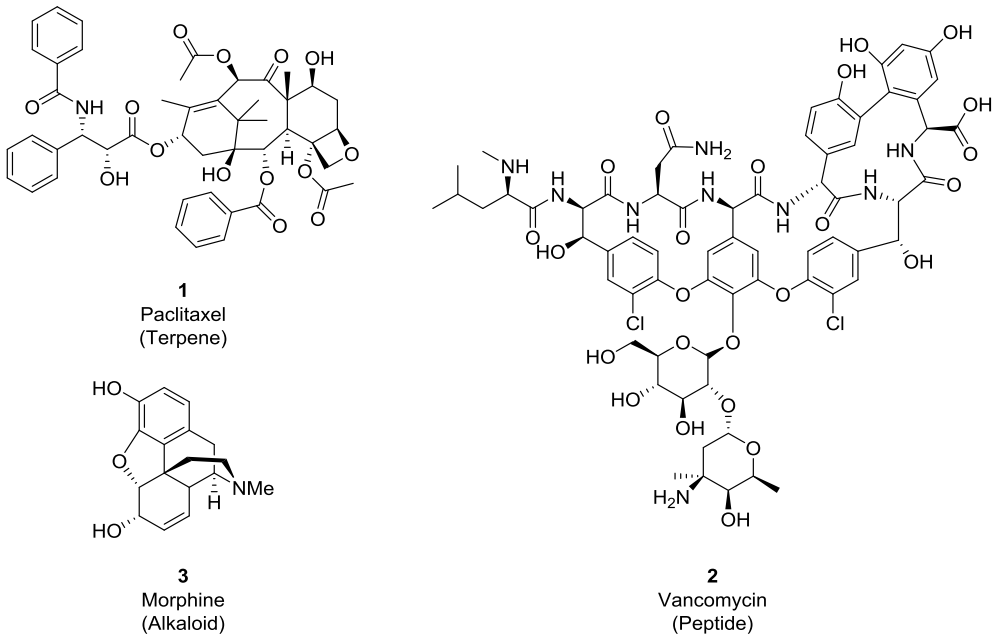


Figure 1 – Examples of natural products.

Polyketides (PK) are another class of natural products which possess wide ranging structural diversity and biological activities.[5] Applications of PK include antibiotic (e.g. erythromycin **4**, Figure 2),[6] antifungal (e.g. natamycin **5**),[7] immunosuppressant (e.g. sirolimus, also known as rapamycin **6**),[8] and anticancer agents (e.g. pederine **7**).[9] Polyketides have been isolated from plants, bacteria, fungi and marine organisms. One such example is the fungal derived polyketide lovastatin **8**, which is used as an anti-cholesterol drug.[10]

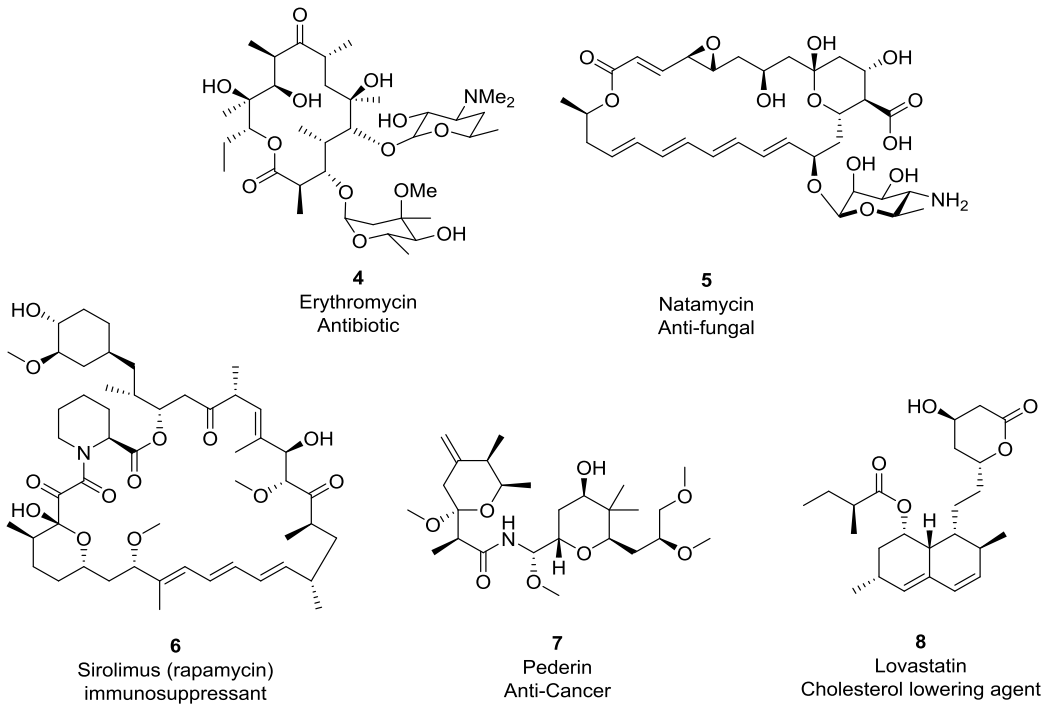


Figure 2 – Examples of bioactive polyketide natural products.

Polyketides are comprised of a carbon backbone which is produced by a class of enzymes known as polyketide synthases (PKS). Once the backbone has been formed, the polyketide can undergo further functionalisation which is carried out by 'tailoring enzymes', further increasing the structural diversity exhibited.

PKS are large enzymes known as megasynthases, which usually comprise over 2000 amino acids and possess multiple catalytic domains.[5] Identifying the genes which encode the PKS and associated tailoring enzymes is relatively simple, as they tend to be clustered together in the genome. The identification of the PKS genes enables the biosynthesis of the polyketide to be studied. Gene knock out and heterologous expression in a host organism allows key intermediates to be isolated and the biosynthetic pathway to be elucidated.[11] It is also possible to produce isolated PKS proteins using gene expression, thus allowing the biochemical selectivity of PKS to be investigated.[12-14]

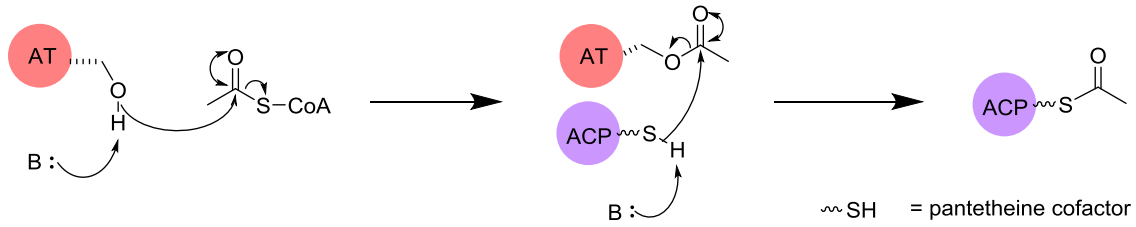
Due to their bioactive properties and interesting structures many polyketides have been targets for chemical total synthesis, which often involves multiple steps and low overall yields due to their chemical complexity. For example the first total synthesis reported for rapamycin **6** involves 102 synthetic steps.[15] By investigating and understanding the biosynthesis of polyketides, it is possible to create mutant PKS which are able to produce drug compounds which are more active and with titres high enough to be suitable for commercial use, for example Wang *et al*. engineered a strain of *Streptomyces avermitilis* resulting in a 300 fold increase in the production of doramectin.[16]

1.2 Fatty acid and polyketide biosynthesis

Although polyketides have great structural diversity, they all share a common biosynthetic origin. Their biosynthesis is closely related to that of fatty acids (FA), the study of which has helped to advance the knowledge of PKS greatly.[5, 17] FA are produced by large megasynthases, known as fatty acid synthases (FAS), which link acetate derived malonyl-CoA units in a head to tail formation, *via* enzymatic reactions to produce a fully saturated alkyl product.[18]

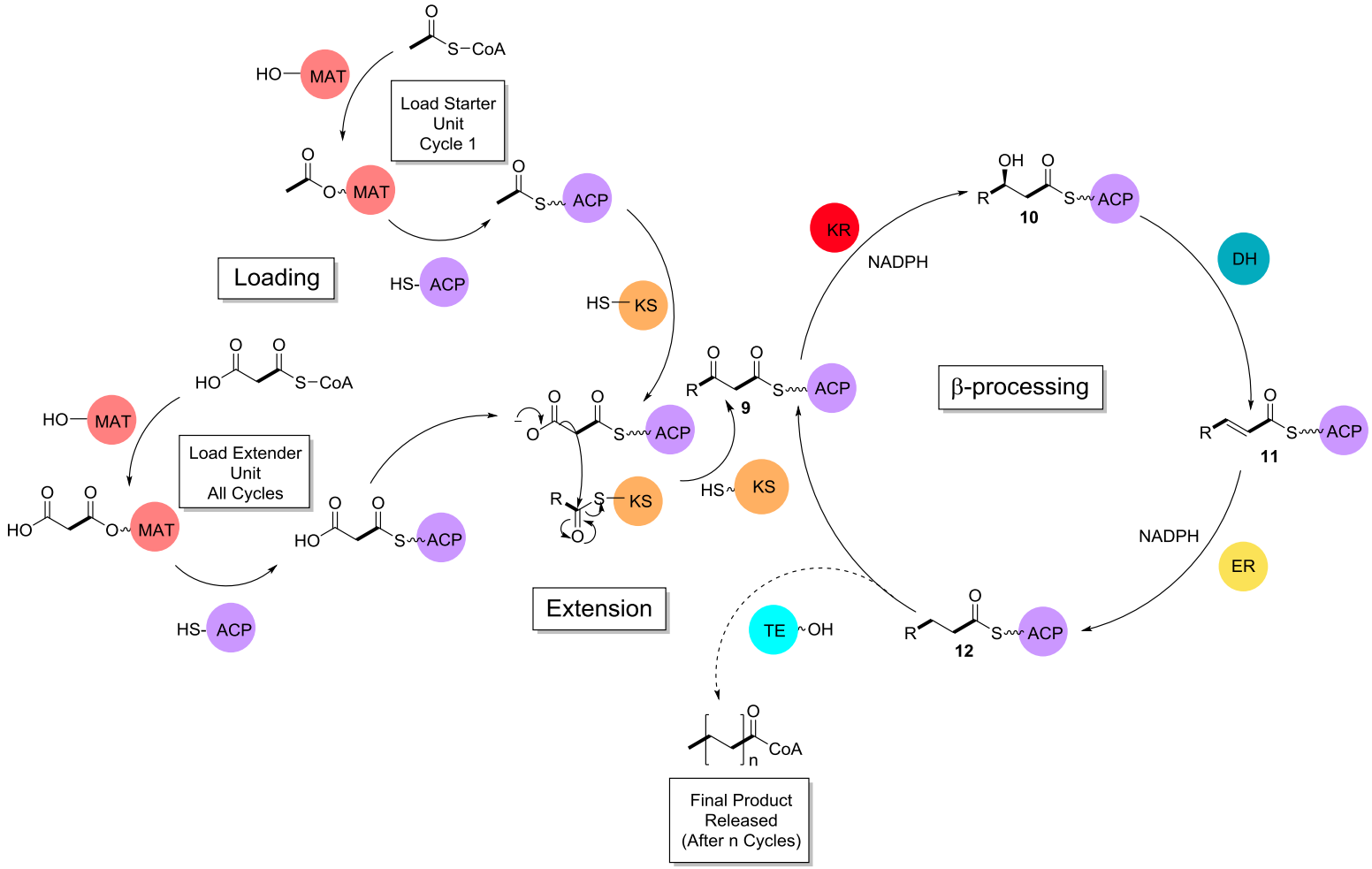
1.2.1 Fatty acid biosynthesis

FA are formed by the linkage of simple building blocks. Straight chain FA utilise acetyl-CoA as a starter unit and malonyl-CoA as an extender unit (Scheme 1).[19, 20] The acetate building blocks are linked head to tail *via* a Claisen-like condensation (Scheme 3).[5, 18] The carbon backbone is then subsequently acted on by various other enzymatic β-processing domains which tailor the backbone. Successive cycles of elongation and β-processing occur until the chain of the required length is produced (Scheme 2). Branched FAs are usually formed by the utilisation of a branched starter unit, for example isobutyl-CoA, isovaleryl-CoA and 2-methylbutyryl-CoA.[20] Malonyl-CoA is still used as the extender unit.

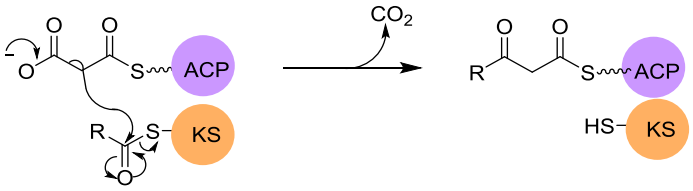


Scheme 1 – Loading of the acetyl-CoA starter unit.

The biosynthesis of FAs can be broken down into two stages; chain extension and β-modification. Chain extension is controlled by three domains: malonyl-acetyl transferase (MAT), acyl carrier protein (ACP), and the β-keto acyl ACP synthase (also known as ketosynthase KS). The first step in the biosynthetic cycle is the loading of the starter and extender units. The MAT binds an acyl Coenzyme A (CoA) starter unit *via* an ester bond, utilising a conserved serine hydroxyl group (Scheme 1).[21] The MAT then transfers the starter unit to the ACP *via* the formation of a thiolester to the terminal thiol of a pantetheine cofactor. The ACP subsequently transfers the starter unit to the KS domain *via* another thiolester to the thiol of a cysteine residue. An extender unit is then loaded in a similar fashion, firstly to AT then transferred to the ACP.



Scheme 2- Fatty acid biosynthetic cycle, malonyl-acetyl transferase (MAT), acyl carrier protein (ACP), ketosynthase (KS), ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), thiolesterase (TE).

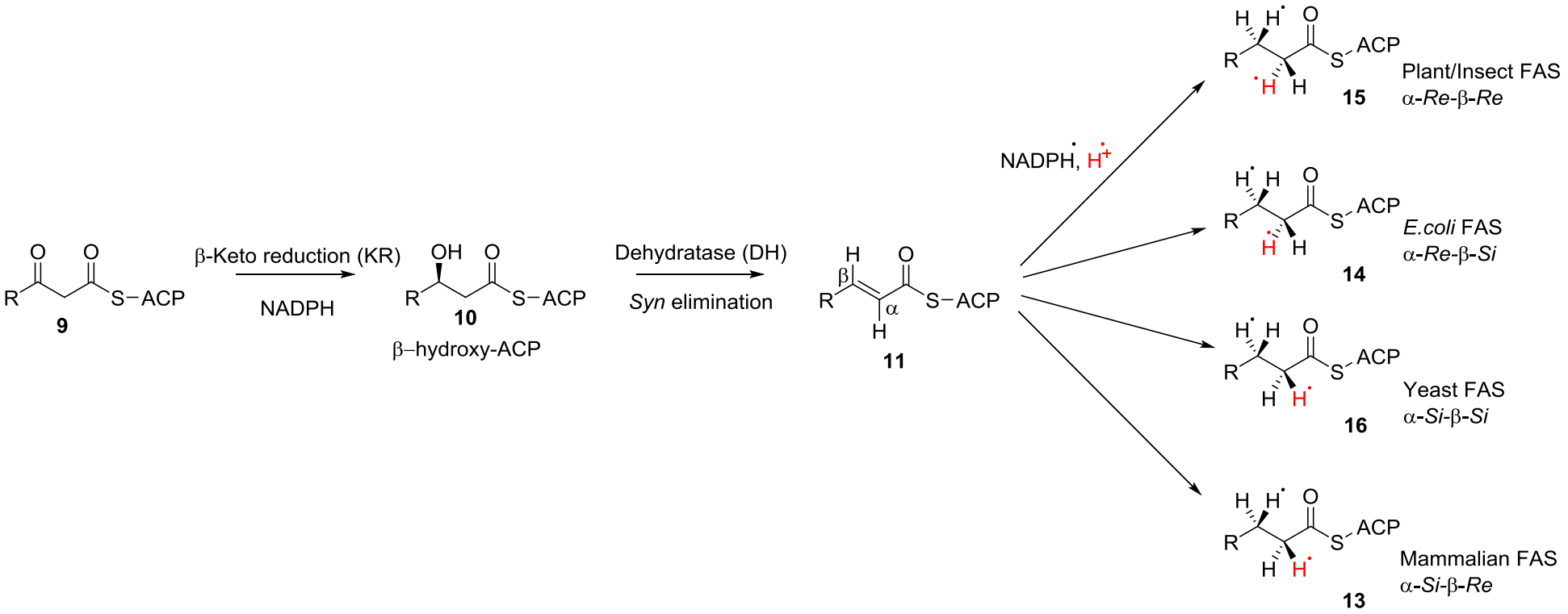


Scheme 3 – Proposed mechanism for the Claisen-like condensation of starter and extender unit in FAS.

Once the β-keto thiolester has been formed it then undergoes a number of β-modification steps. These catalytic reactions are controlled by the ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) enzymes (Scheme 2). The KR domain catalyses the reduction of the β-keto thiolester **9** to a β hydroxyl group using NADPH as a co-factor.[22] The reductions result in the formation of β-hydroxyl-acyl ACP **10**.[23] The alcohol can then be dehydrated by the DH domain to give an α,β-unsaturated thiolester **11**.[22] The dehydration reaction occurs *via* a *syn* elimination mechanism, where the 3-hydroxy and 2-hydrogen are lost from the same face. This results in an *E* alkene, forming crotonyl-ACP **11** (Scheme 4).[24]

The ER domain reduces the enoyl ACP **11** to give a fully saturated alkyl chain **12**, using NADPH as a cofactor. The stereochemical course of the enoyl reduction varies from organism to organism (Scheme 4). The hydrogens can be added to either the *Re* or *Si* face of the alkene. In rat FAS the hydrides are added to the α-*Si*-β-*Re* faces of the alkene **13**.[25] In *E. coli* FAS the hydrides are added to the α-*Re*-β-*Si* faces **14**, this is the opposite to the addition observed in mammalian FAS.[26] In plant and insect FAS it was found that the addition of the hydride occurred at the α-*Re*-β-*Re* faces **15**,[27] where as in yeast the addition happened at the α-*Si*-β-*Si* faces **16**.[23]

The fatty acid chain is then ready to undergo another round of chain elongation and β-processing. The cycle is repeated until the alkyl chain has reached the desired length, then a release domain such as a thiolesterase (TE) acts by catalysing the release of the final molecule (Scheme **2**).[28]



Scheme 4 – Stereochemical outcome of fatty acid biosynthesis, showing the possible modes of H addition during enoyl reduction.

1.2.2 Fatty acid synthase classification

All FAS contain the same five domains, ACP, KS, KR, ER and DH. Other domains present vary between species of FAS. Mammalian contains an embedded thiolesterase (TE) domain as well as a malonyl-acetyl transferase (MAT).[21, 29] Yeast FAS possesses a number of unique features. Instead of a MAT, it has an acetyl transferase (AT) which is dedicated to priming the FAS with the acetyl-CoA starter unit. The extender units are loaded using a malonyl-palmitoyl transacylase (MPT) which also off-loads palmitoyl products.[29] Yeast FAS also does not contain a TE, instead the finished fatty acid is transferred to the MPT, where the fatty acid is transferred to CoA and released.

Although all FAS comprise the same key enzymatic domains, the domain organisation can differ, which gives rise to different classifications of FAS (Figure 3). In Type I FAS the enzymatic domains are covalently linked and organised into a large multi-functional complex.[30-32] Type I FAS are most commonly found in animals and fungi. Type II systems consist of individual enzymes and are typically found in bacteria.

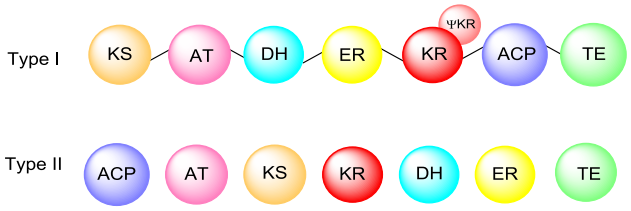


Figure 3 – Cartoon of FAS domain organisation, type I the enzymatic domains are covalently link, type II the enzymatic domain are discrete proteins.

Although it is possible to classify FAS by type, their structural organisation can vary significantly. For example type I yeast FAS has a different structure to that found in mammalian FAS. Type I yeast FAS forms a barrel-like structure which comprises six "reaction chambers".[29] Each reaction chamber contains one complete set of catalytic domains. The barrel is made of two dome shaped units, with each unit containing different catalytic domains. The α-unit contains four catalytic domains (KS, KR, ACP, and PPT) and two structural units. The β-unit comprises four catalytic domains (AT, ER, DH, and MPT) and four structural units. Mammalian type I FAS, on the other hand, adopts an X-shape, with a central body with extended arm and leg-like structures on the upper and lower ends (Figure 4).[33]

The X-shape is divided into two units horizontally across the middle. The structure is discussed in more detail in Section 1.3. The lower unit contains two catalytic domains (KS, MAT) and no additional structural domains. The upper unit comprises three catalytic domains (DH, ER and the KR). FAS also contains two additional domains, the ACP and TE domains. Both of these domains are not visible in the porcine FAS crystal structure published by Maier *et al.* due to

their inherent flexibility.[34] It is thought that the ACP and the TE domains are located on the periphery of the upper or lower unit, but this is yet to be confirmed by crystal structure.

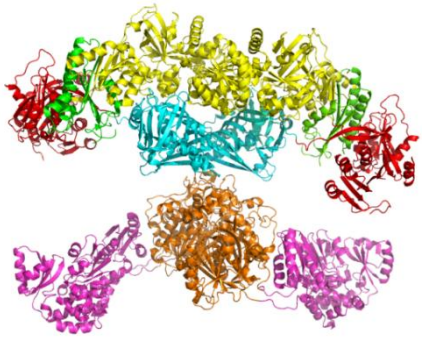


Figure 4 - Crystal structure of FAS (PDB: 2VZ8).

In Type II FAS, each enzymatic domain is expressed as a discrete protein, which can be isolated individually. These are typically found in plants, bacteria and algae. [30-32, 35] It is thought that Type II FAS, although discrete proteins, form a highly organised complex.[36]

1.2.3 Polyketide biosynthesis

There is a high degree of chemical and structural homology between FAS and PKS. This homology results in the biosynthesis of polyketides being similar to that of fatty acids. As with FAS, simple building blocks are condensed together to form the backbone of the molecule. PKS are able to utilise a wide range of starter units (Figure 5).

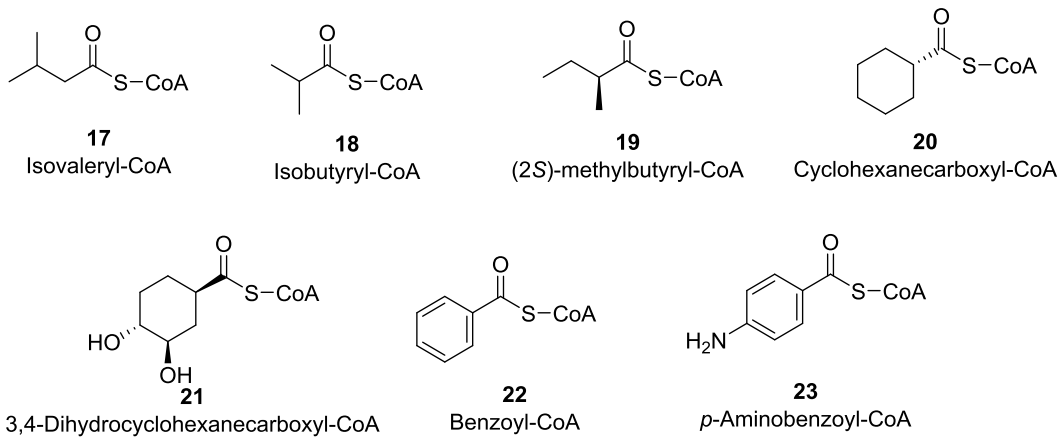


Figure 5 – Starter units which can be utilised during polyketide acid biosynthesis.

Type I PKS commonly utilise acetate and propionate derived starter units.[37] Type I PKS have also been known to utilise more novel starter units. For example isovaleryl-CoA **17** is used as the initial building block for myxothiaxol,[37] whereas isobutyryl-CoA **18** and (*S*)-2-methylbutyryl-CoA **19** are the initial building blocks during the synthesis of avermectins.[38] Cyclohexanecarboxylate (CHC) **20**,[39] 3,4-dihydroxycyclohexanecarboxylate (DHCHC) **21**,[40] benzoate **22**[37] and *p*-aminobenzoate **23** starter units have also been observed.[41]

PKS can also utilise a wide range of extender units (Figure 6). A typical extender unit is malonyl-CoA **24**,[42] but branched extender units such as 2-methylmalonyl-CoA **25** and 2-ethylmalonyl-CoA **26** are also used. Extender units containing halogen atoms, **27**, ethers **28**, hydroxyl groups **29** and amines **30** have also been observed.

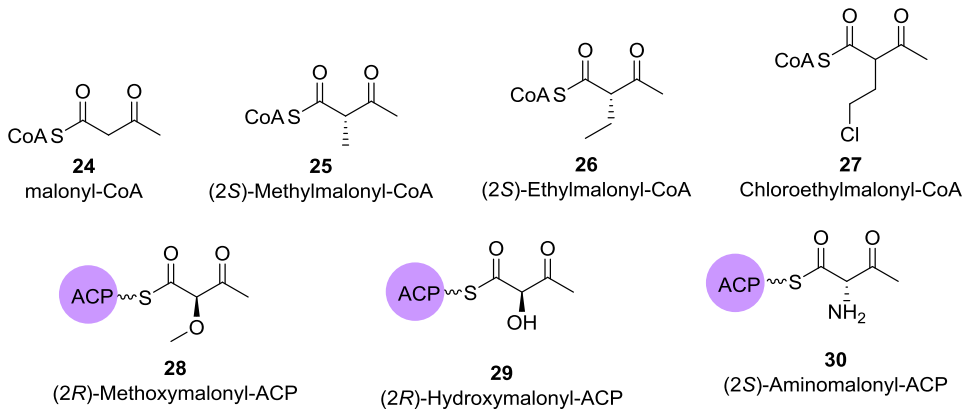
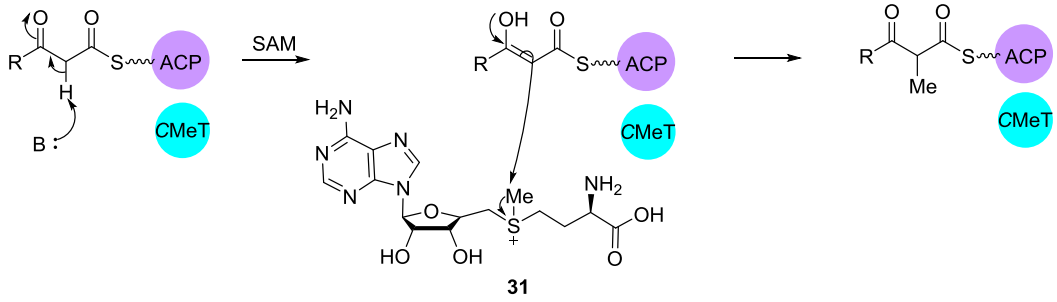


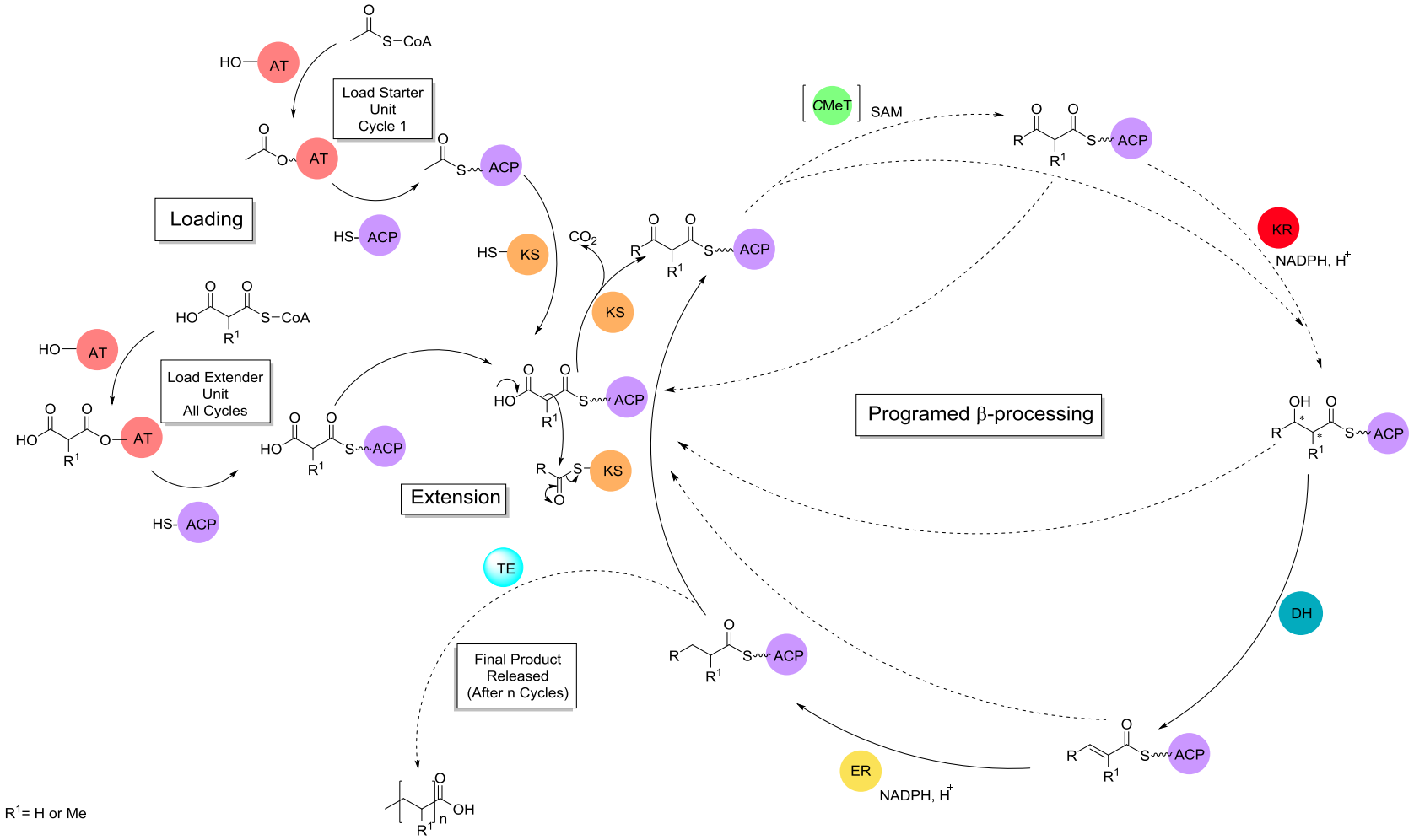
Figure 6 - Extender units which can be utilised during polyketide biosynthesis.

PKS join C_2 starter and extender units together to form the backbone of the chain utilising the AT, ACP and the KS domains. PKS are then able to tailor the backbone of the chain using the same enzymatic domains as the FAS (KR, DH and ER). Some PKS are able to carry out an additional tailoring reaction, such as methylation at the α-carbon. Methylation reactions are catalysed by a *C*-methyltransferase (*C*-MeT) domain, using *S*-adenosyl methionine (SAM, **31**, Scheme 5) as a cofactor. Methylation occurs before any of the other tailoring reactions take place. While both mFAS and fungal PKS contain a *C*-MeT domain, in FAS this domain is inactive. Many fungal type I PKS on the other hand contain an active form on this domain.



Scheme 5 - *C*-MeT domain catalyses transfer of a methyl group from *S*-adenosyl methionine to the β-keto thiolester.

In addition to controlling the number of chain extensions which occur, PKS have additional levels of programming.[5] PKS have the ability to control starter and extender selection. PKS are able to control which domains are active in each catalytic cycle, allowing the extent of reduction and dehydration in each round of extension to be controlled, this results in a variety of functional groups being formed (Scheme 6).



Scheme 6 – Polyketide biosynthetic cycle. The methyl group can arise from the incorporation of a methylated starter or extender unit; in fungal systems an active *C*Met domain carries out a methylation reaction at the α-positions.

1.2.4 PKS classification

PKS are categorised into three broad categories, type I, II and III.[5, 9, 18] Type I and II classifications arise from their similarities to fatty acid biosynthesis. Type I PKS comprise multiple catalytic domains covalently linked together to form large multifunctional proteins. Type II PKS are formed of discrete proteins which each catalyse a specific reaction. Type II and III PKS both produce aromatic compounds but do so in a fundamentally different way to one another.[32]

1.2.5 Type I PKS

Type I PKS are the most complicated and the most versatile of all the PKS catalysts. They are further subdivided into two categories, modular and iterative.[9] Type I Modular Polyketide Synthase (mPKS) contain more than one distinct module of enzymatic domains.[9] There is a distinct module for loading and for every subsequent catalytic cycle, with every catalytic domain within each module only acting once. The presence or absence of different β-processing domains dictates the structure of the final product, such as the ER or DH domain. Typically type I mPKS produce reduced compounds. 6-Deoxyerythronolide B Synthase (DEBS) is the type I mPKS responsible for producing the macrolide ring precursor of erythromycin **4** (Figure 7).

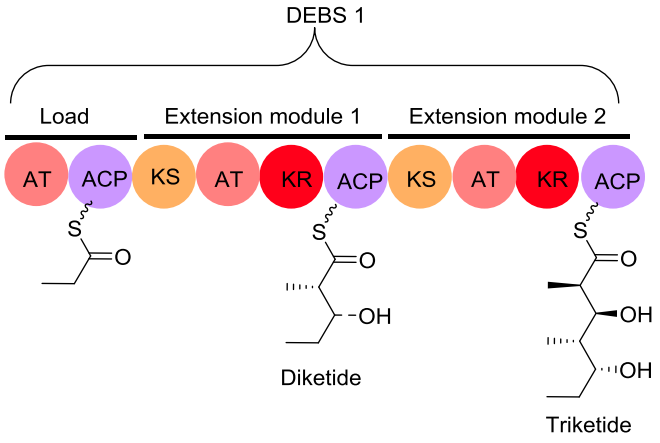
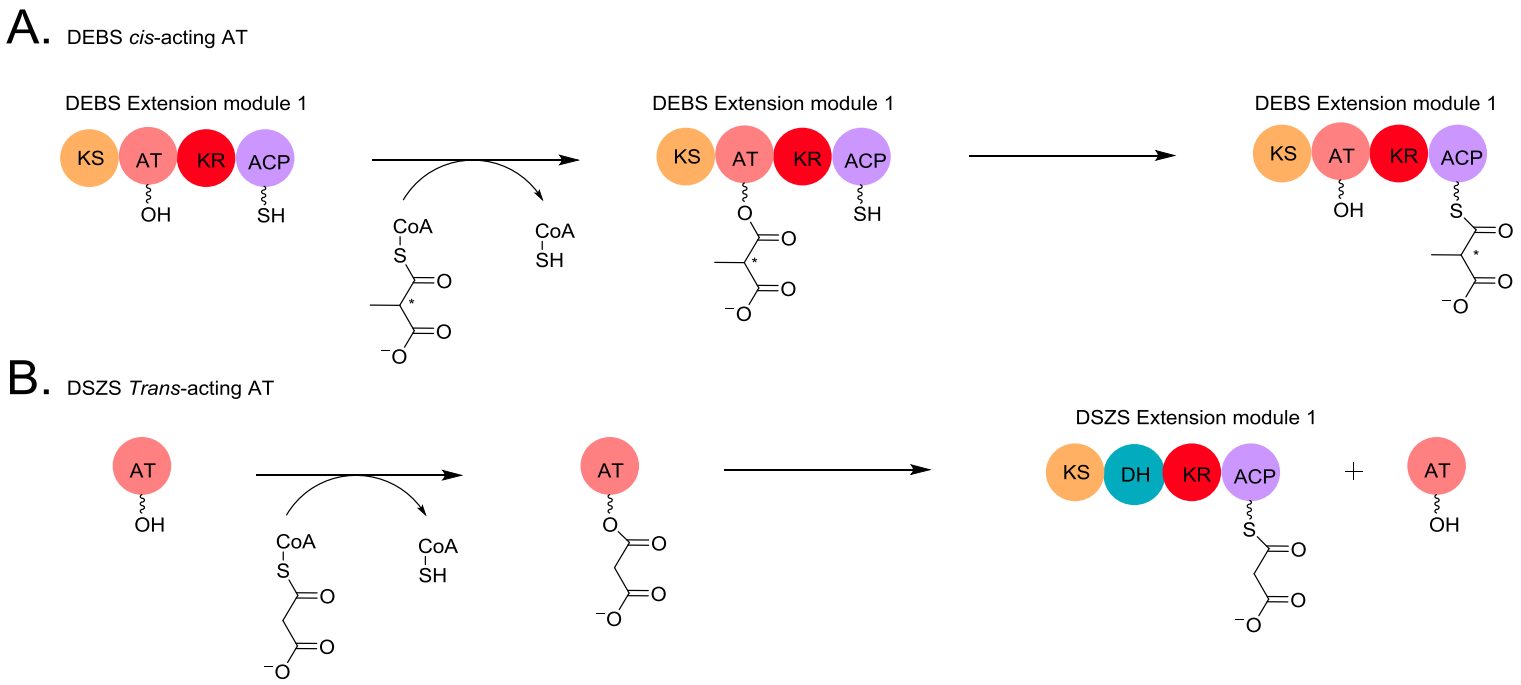


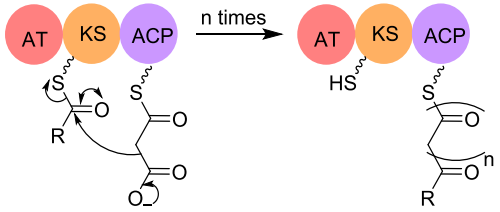
Figure 7 - Illustration of programming in DEBS mPKS. The loading module and the first two extension/modification modules of erythromycin PKS shown.

There are three core enzymatic domains required for chain extension AT, KS and ACP. Type I mPKS are known to contain two types of AT.[43] If the AT domain is paired with its own ACP, it is referred to as a *cis*-acting AT (Scheme 7, A). DEBS is an example of a PKS with *cis*-acting AT domains, each module contains an AT and it is paired to an ACP. *Trans*-acting AT are stand alone enzymes which can work iteratively to transfer extender units to one or more of the ACP in the PKS assembly line. The DSZS PKS responsible for the production of dizorazole is of the *trans*-acting AT type (Scheme **7**, B).



Scheme 7 – The different modes of *cis* and *trans*-acting AT: **A,** shows the *cis*-acting AT of DEBS loads the extender unit from its corresponding CoA to the ACP within the same module; **B**, shows the extender unit for DSZS is first loaded onto *trans*-acting AT and subsequently loaded onto the ACP in module 1 of the DSZS PKS.

Unlike mPKS, Type I iterative PKS (iPKS) consist of a single modular architecture which is used repeatedly (Scheme 8).[9] The single module contains all the domains required to produce the compound. Each catalytic domain acts repeatedly to extend and tailor the growing polyketide chain. iPKS can produce both reduced and aromatic compounds. The iPKS undergoes a defined number of chain extensions in order to produce a polyketide chain of the desired length. Norsolorinic acid (NSA) is an example of a iPKS which contains only a KS, AT and an ACP (Scheme 8).[44]



Scheme 8 - A "minimal" iPKS, which contains the 3 key domains, AT, KS, ACP.

The type I iPKS found within fungi are further subdivided into non-reducing (NR), partially reducing (PR) and highly reducing (HR) types.[9, 44] This classification arises due to the incorporation of certain β-processing catalytic domains, ketoreductase (KR), dehydratase (DH) and enoylreductase (ER).

Non-reducing iterative polyketide synthases (NR iPKS) produce compounds which require no reductions during chain extension, for example citrinin **32** (Figure 8). [5, 9] The synthase contains a minimal number of domains (AT, KS, ACP), but it lacks all β-keto processing domains (KR, DH, ER).[5, 9] NR PKS domains are typically organised as follows; a starter unit loading domain (SAT), the KS and an AT for loading extension units. After the AT domain is a conserved domain, (PT, which helps control chain length of the product and act as a product template domain). The final domain is the ACP, some NR iPKS terminate after the ACP. Many have been found to have other domains after the ACP such as *C*-methyltransferases and reductases involved in product release (Figure 8).

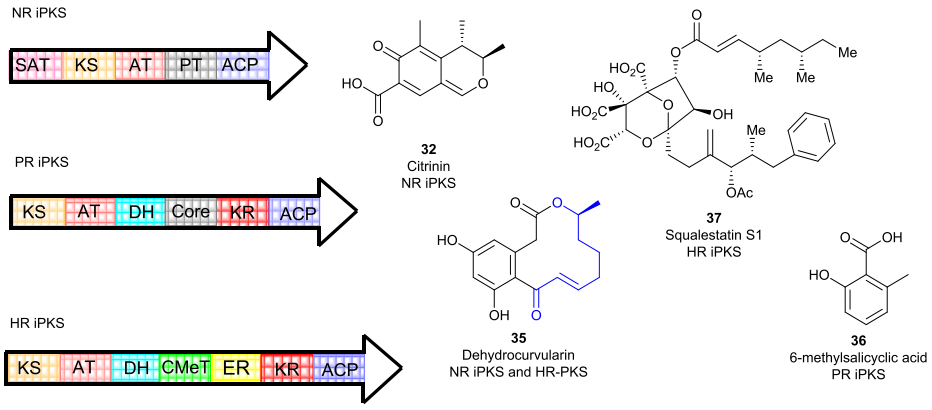


Figure 8 – General domain architecture of NR iPKS, PR iPKS and HR iPKS, with example compounds produced by the PKS.

Studies of the genes involved in non-reduced polyketide compounds have found that the NR iPKS can incorporate advanced starter units, which have been derived from dedicated FAS, another PKS or an acyl-CoA. Monocerin **33** and zearalenone **34** are examples of products from NR PKS which incorporate advanced starter units (Figure 9). The structures of these compounds suggest high levels of reduction occurs in the early stages of biosynthesis.[5] The later stages of biosynthesis indicate that no reduction reactions occur. Labelling studies have shown that the NR PKS responsible for the production of dehydrocurvularin **35** utilises a tetraketide starter unit formed by a separate HR-PKS (Figure 8).[5]

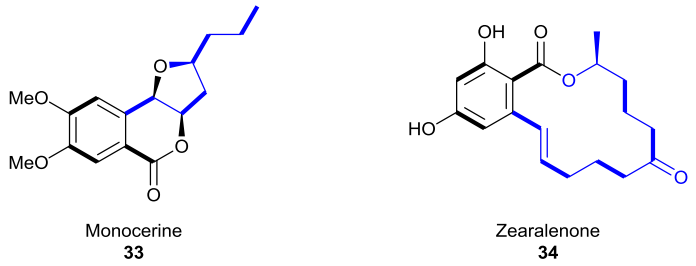


Figure 9 – Examples of NR PKS which potentially incorporate advanced starter units (shown in blue).

Partially reducing (PR) iPKS have a domain structure more closely related to FAS and contain some β-keto processing domains.[5] Typically they possess a KS, AT, DH, KR and an ACP domain. In addition they possess a "core domain", which is located between the DH and KR domains. It is thought that the core domain is responsible for controlling the formation of the synthase. The core domain forms the motif required for key subunit-subunit interactions.[5] Unlike NR iPKS, they do not possess a SAT or PT domain. A well-studied example of a PR iPKS product is the tetraketide 6-methylsalicylic acid, **36**, which is formed by the MSAS PKS (Figure 8).[18]

Highly reducing (HR) iPKS produce complex highly reduced compounds by carrying out a range of β-keto processing reactions.[5] HR iPKS contain the KS, AT and the ACP required for chain elongation. In addition they contain a KR and DH. It has been found that many fungal HR iPKS contain a *C*MeT domain, some also contain an ER domain (Figure 8).[5] Lovastatin **8**, is a well-known product of a HR iPKS.[10] It is formed of two polyketides; a nonaketide, produced by lovastatin nonaketide synthase (LNKS) and a diketide, produced by lovastatin diketide synthase (LDKS). Another well-known compound produced by a HR iPKS is squalestatin S1 **37 (**Figure 8).[5]

A large number of bioactive polyketide derived compounds have been found which also contain amino acid moieties. Fusarin C, **38** consists of a heptaketide fused to a homoserine,[5, 45] Tenellin, **39**, utilises a rearranged tyrosine moiety to form the pyridone ring (Figure 10).[45, 46]

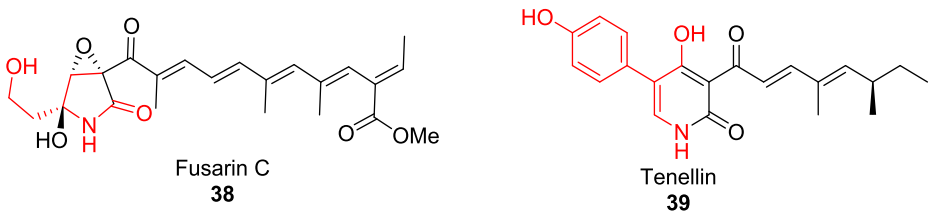
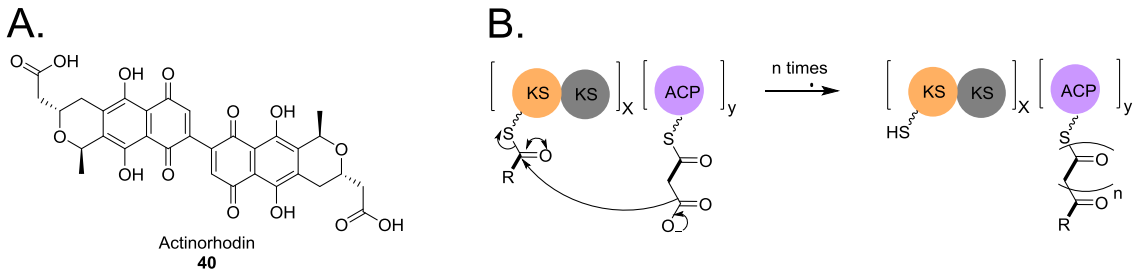


Figure 10 – Example compounds produced by HR PKS-NRPS, with the amino acid derived moieties highlighted in red.

All of these compounds are produced by hybrid PKS-NRPS systems. A non-ribosomal peptide synthase (NRPS) downstream of a HR PKS is responsible for adding the amino acid derivatives which are then incorporated into the PKS structure. The PKS and the NRPS can be combined in two ways. In modular PKS-NRPS, the PKS modules and the NRPS modules are organised in a processive production line. In iterative systems, an iPKS is followed by a single NRPS module; this domain organisation is found predominantly in fungi.[45]

1.2.6 Type II PKS

Type II PKS are composed of discrete proteins which associate together to form an active complex. Actinorhodin **40**, is a blue pigmented antibiotic which is produced by a type II PKS (Scheme 9, A).[47] The biosynthesis of **40** has been extensively studied and is thought of as a model compound for studying the type II PKS, both the key enzymes and the subsequent post-PKS tailoring enzymes.

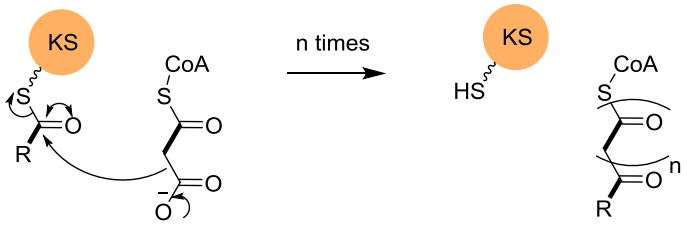


Scheme 9 – **A,** Structure of actinorhodin; **B**, Representation of a type II PKS, the catalytic KS_{α}, is shown in yellow and the structural KS_{β} is shown in grey.

The complex works in an iterative manner, typically producing aromatic products.[9] A minimal PKS, which contains two KS domains (one catalytic, KS_{α}, the other structural, KS_{β}) carries out the repeated condensation reaction. The growing polyketide chain is attached to an ACP (Scheme **9**, B). Typically an acetate starter unit is used to initiate the cycle, with malonate as the extender unit. A malonyl-CoA ACP (MCAT) recruits extender units. Other proteins can co-operate with the minimal PKS to tailor the chain to the desired product, which can include KR, cyclases (CYC) and aromatase (ARO) domains.

1.2.7 Type III PKS

Type III PKS are not organised into a multi-enzyme complex like type II PKS, instead a single KS-like active site repeatedly catalyses the condensation of acetate units (Scheme 10).[9] Like type II, the type III PKS typically produce aromatic compounds which are mono or bicyclic in nature. Unlike type I and II PKS, type III PKS do not utilise an ACP to carry the growing polyketide chain, instead the condensations are activated by a derivatised CoA starter unit.[9]



Scheme 10 - Representation of a type III PKS.

1.3 Structure and composition of FAS and iPKS proteins

1.3.1 Structure of mFAS

The crystal structure of a mammalian type 1 FAS (mFAS) has greatly increased our knowledge of the structure of mFAS (Figure 11).[33, 48] mFAS consists of a 270 kDa polypeptide chain. In its active form mFAS assembles into an intertwined homodimer with an approximate "X" shape. The catalytic domains are organised into two units. The lower unit contains the condensation domains, ketosynthase (KS), and the malonyl-acetyl transferase (MAT). The upper unit contains the modification domains, the KR, DH and the ER. In addition to these three main β-modification domains, mFAS possesses two additional non-modification domains located on the periphery of the active domains.[35, 48]

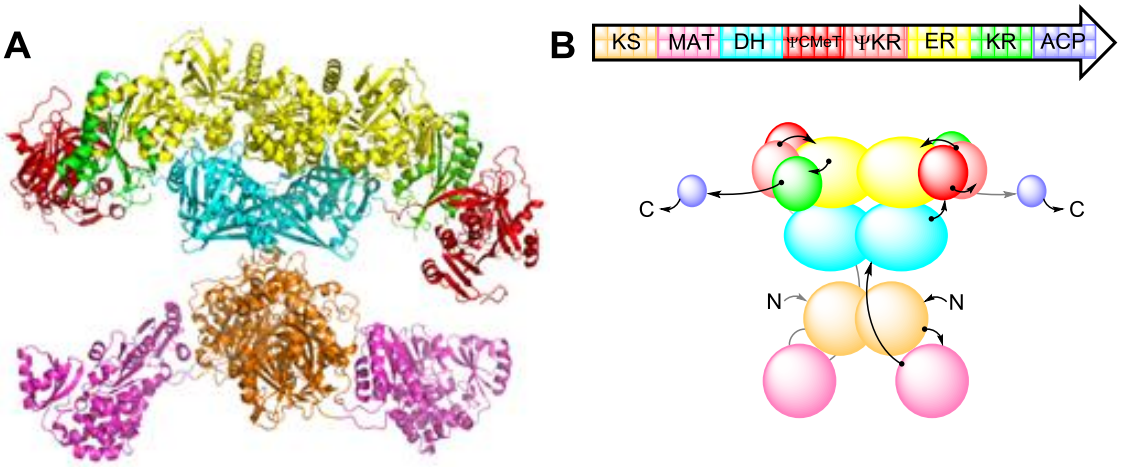


Figure 11 - Crystal structure of FAS (PDB: 2VZ8) and Cartoon structure of mFAS, the *C*-MeT domain (red) is present, but inactive.

The first of the non-modification domains is the pseudo-methyltransferase (ψ*C*MeT), a domain homologous to the *C*-methyltransferase family in structure but having low sequence identity.[48] In order for the *C*MeT to be active it must be able to bind a SAM (**31)** co-factor. In mFAS the SAM (**31)** binding motif, ExGxGxG has mutated to ExLxGxG, resulting in the loss of co-factor binding site and subsequent loss of activity.[34, 48]

The second non-modification domain is a structural part of the KR, referred to as the pseudo ketoreductase (ψKR). The dimerisation interface between the ψKR and the KR provides structural support to the active site of the catalytic KR.[48]

1.3.2 Dimer interface

The characteristics of a multi-enzyme complex are dictated by the nature of the interactions and linkers between the individual domains. Only 9% of the mFAS sequence is responsible for linkers between catalytic domains.[48] The structural organisation of the complex is different to the organisation of the linear protein sequence. Very few of these linkers are present for structural integrity. There is only one structural linker responsible for connecting MAT to the KS in the condensation block. There is no direct contact between the MAT and KS domains.[48]

The two polypeptides dimerise due to homophilic interactions of certain contact areas. The ER and KS form the main area of contact between the two polypeptide chains. The DH forms an additional area of contact between the two peptide chains. Within the β-modification unit, the KR is the central connector to all the other modification domains, interacting with the DH, ER and non-catalytic ψKR and ψ*C*Met. Neither the DH nor the ER directly interact with the non-catalytic domains.

1.3.3 Relation of mFAS to PKS structure

Although currently there are no intact crystal structures of full length iPKS or mPKS, isolated PKS domains have been crystallized.[6, 17, 49-52] The structural insights gained from these crystallized domains has provided evidence of the evolutionary relationship between mFAS and HR-PKS.[35, 48] On a sequence level, mFAS domains have a high sequence identity to HR-PKS domains.[32] mFAS is more closely related to HR-PKS than to bacterial FAS. The similarities between mFAS and HR-PKS extend beyond the sequence level, to the linkers, the domain order and the architecture of certain domains.[48]

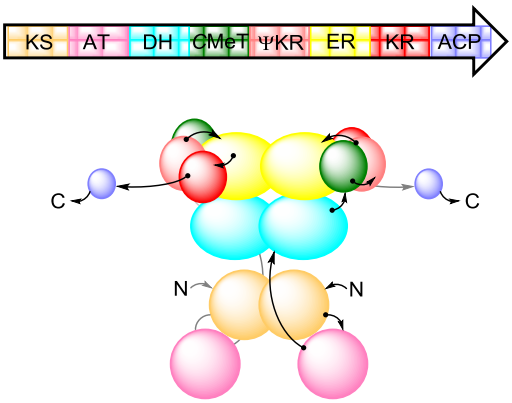


Figure 12 - Cartoon structure of PKS, The *C*-Met domain (green) is present and active.

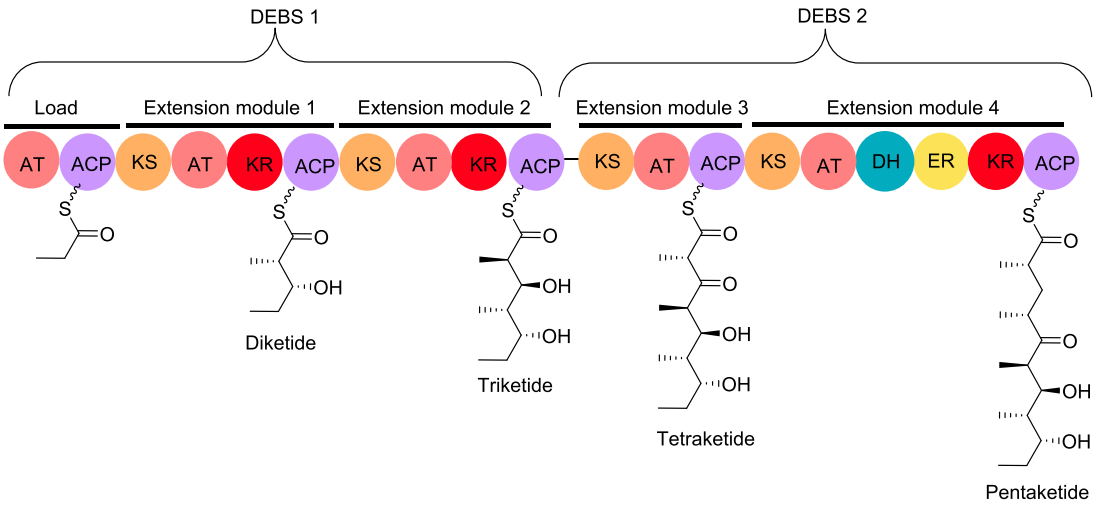
The similarities in protein sequence between mFAS and PKS suggests that the domain organisation within the PKS is likely to be similar to that found in mFAS (Figure 12).[17] Three crystal structures which were published almost simultaneously highlight the clear evolutionary connection between type 1 FAS and PKS.[53] First, the crystal structure of porcine FAS was published by Ban *et al*.[34] Although the resolution of the structure was not high enough to allow detailed amino acid mapping it was possible to fit previously reported domain structures with a high degree of confidence. The discovery of the structural KR domain by Keating Clay *et al.* resulted in the assignment of a large segment of linker sequence across the whole family of type I PKS modules.[54] FAS contains an analogous region of linkers, which is highly likely to correspond to a structural KR domain. If the organisation of the type I FAS and PKS KR domains is similar it can be hypothesised that this commonality can be extended to the dimeric organisation of the DH and ER domains.[53] Finally the crystal structure of the erythromycin KS-AT region published by Khosla *et al.* showed striking similarities to lower half of type I FAS, further strengthening the connection between the FAS and PKS architecture.[55]

mPKS function in a different way to mFAS, with each PKS module catalyzing only a single round of extension and β-modification. Each mPKS module contains the loading and condensation domains as well as a variety of β-carbon processing domains. mPKS modification units can have domains which are inactive and in some cases absent.[9] For example DEBS module 1 only possesses one β-carbon modification domain, KR-ψKR.[48] Comparisons of the KR-ψKR in DEBS and of the corresponding Section of mFAS found that the insertion of the ER domain in mFAS had no effect on the core folds of the KR-ψKR. The structural similarities found between mFAS and PKS imply that iterative and non-iterative chain elongation can be achieved using modules with overall similar architecture.

1.4 Programming of PKS

Compared to fatty acids, Polyketides display a very wide diversity of chain-lengths and functionalisation. PKS must, therefore, have the ability to precisely control chain-length and the cycle of β-modifying reactions. They also have diverse alkyl substituents, added by methylation or use of varied extender units. This ability to control all these aspects of biosynthesis is referred to as *programming*.[5]

In mPKS, each chain extension is catalysed by a discreet module containing only the required catalytic domains (Scheme 11). The programming within the PKS is explicit in the number of modules (which controls chain length) and the composition of those module's β-processing domains.



Scheme 11 – The first four modules in the DEBS PKS.

The programming within iPKS is far more subtle and less well understood. iPKS comprise a single module containing one copy of each catalytic domain, although not all the domains are active in every round of chain extension.[9] Many questions surround the programming of iPKS, such as: how is the selection of starter and extender units controlled, what controls the length and release of the product; and also what controls the degree of β-processing.

Work by Townsend and co-workers investigating PksA, an iPKS, which is responsible for the initial stages of the biosynthesis of aflatoxin B_1, has given an insight into how these processes are controlled in NR PKS.[56-59] Their work identified that the starter unit acyl transferase (SAT) controls the selection of starter unit,[56] whereas the selection of the extension unit was controlled by the malonyl-acetyl transferase (MAT).[56, 57] It was also found that while the product template (PT) domain controlled chain length and cyclisation and aromatisation,[58] the SAT, MAT, KS and ACP domains also helped control the chain length.

The programming of HR-PKS is more complex as they do not possess SAT or PT domains and the presence of *C*MeT, KR, ER and DH domains must contribute to the programming. It is

possible that individual catalytic domains possess selectivity for specific substrates. For example, the chain length could be controlled by the KS or AT recognising the length of the growing acyl-ACP. The AT domain could recognise the full length chain and not transfer it to the KS for further chain extension. Another possibility is that once the chain has reached the desired length the KS no longer recognises it and subsequently cannot catalyse the chain extension. The methylation pattern of a substrate may also play a part in whether or not a substrate is recognised.

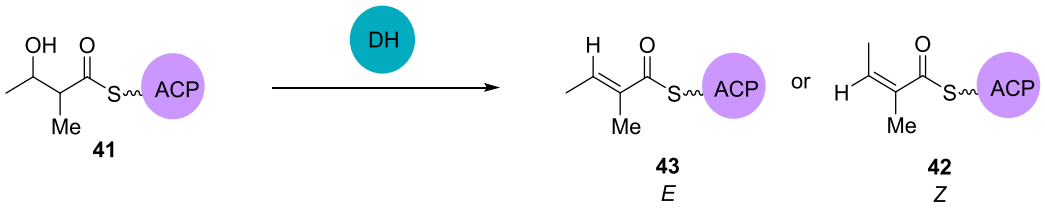
Work by Fisch *et al.* investigated the programming of the HR-PKS involved in the biosynthesis of tenellin **39** and desmethylbassianin.[45] Catalytic domains from the tenellin PKS (TENS) and desmethylbassianin PKS (DMBS) were systematically swapped and the chain length and methylation patterns observed. The products formed gave an insight into the programming of these two PKS. It was found that neither the TENS nor DMBS KS, AT or ACP domains exerted any control over the substrate chain length and will indiscriminately extend the acyl ACP.

The *C*Met domain on the other hand exhibited a high level of programming, the DMBS *C*Met was found to only recognise a diketide substrate, whereas the TENS *C*Met domain recognised both diketide and triketide substrates. The KR domain also exhibited programming which controls the chain length of the product. The TENS KR was able to act on di-, tri- and tetraketides but not pentaketides. In contrast, the DMBS KR was able to act on all substrates up to the pentaketide but not hexaketide. It was proposed that when the KR domain fails to act the chain is released.

1.5 DH domains from FAS and PKS

1.5.1 Function

The dehydratase domain catalyses the dehydration of the β-hydroxy acyl ACP **41** to an α,β-unsaturated thiolester (Scheme 12).[17, 21, 60] The resulting α,β-alkene can potentially take either *Z* **42** or *E* **43** geometry (Scheme 12).



Scheme 12 - Function of the DH domain and the possible geometry of α,β-unsaturated thiolester.

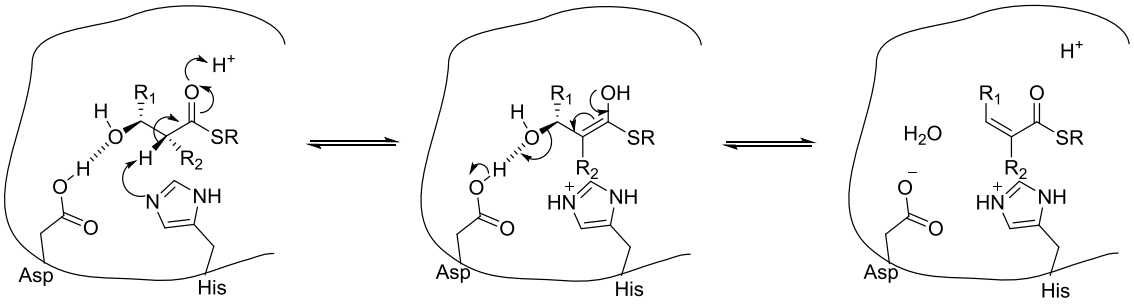
A great deal of our understanding of how the DH domain functions comes from the study of FAS and mPKS.[61] There have been no reports of investigations of the workings of the iPKS DH domain. Sequence alignment of mFAS and mPKS shows limited sequence identity (Figure 13), mFAS is 18% to 23% identical to crystallised mPKS domains. Despite the low sequence

identity, analysis of the crystal structures has found mFAS and PKS DH domains have many structural features in common and these will be discussed in more detail in Section 1.5.3.

```
#  Percent Identity  Matrix - created by Clustal2.1

     1: SQTKS_DH      100.00   12.69   17.96   16.91   17.27   18.06   19.42   18.18
     2: 2vz8_mFASDH    12.69  100.00   22.79   21.48   18.18   15.81   21.67   20.08
     3: 4LN9_Rif10DH   17.96   22.79  100.00   43.83   21.77   23.67   26.01   27.11
     4: 3EL6_Ery4DH    16.91   21.48   43.83  100.00   18.56   21.22   21.80   21.66
     5: 3KG7_CurH_DH   17.27   18.18   21.77   18.56  100.00   26.06   23.97   23.02
     6: 3KG8_CurJ_DH   18.06   15.81   23.67   21.22   26.06  100.00   28.67   29.31
     7: 3KG6_CurF_DH   19.42   21.67   26.01   21.80   23.97   28.67  100.00   68.42
     8: 3KG9_CurK_DH   18.18   20.08   27.11   21.66   23.02   29.31   68.42  100.00
```

Figure 13 – Percentage identity matrix of mFAS (2), crystallised mPKS DH domains (Rifamycin DH, Rif10DH (3), Erythromycin DH, Ery4DH (4), Curcin DH, CurH (5), CurJ (6), CurF (7), CurK (8)) and SQTKS (1), identified by PDB numbers, full alignment in Appendix 1.

The active site of PKS and FAS DH domains consists of a histidine (H) and an aspartic acid (D), which are highly conserved in both FAS and PKS.[6, 12, 13, 17, 48, 60-64] Mutagenesis of these amino acids has demonstrated their catalytic importance.[60, 65] The mutation of catalytic H in borrelidin BorDH2[60] and pikromycin PikDH2 to a phenylalanine (F) and alanine (A), respectively, resulted in total loss of domain activity, showing that the H is vital to catalytic activity.[65] The catalytic H abstracts the α-proton and the D donates a proton to the leaving β-hydroxyl group.[6] The proton and water are lost from the α and β-carbon positions. The mechanism is most likely to be E1cb (Scheme 13).



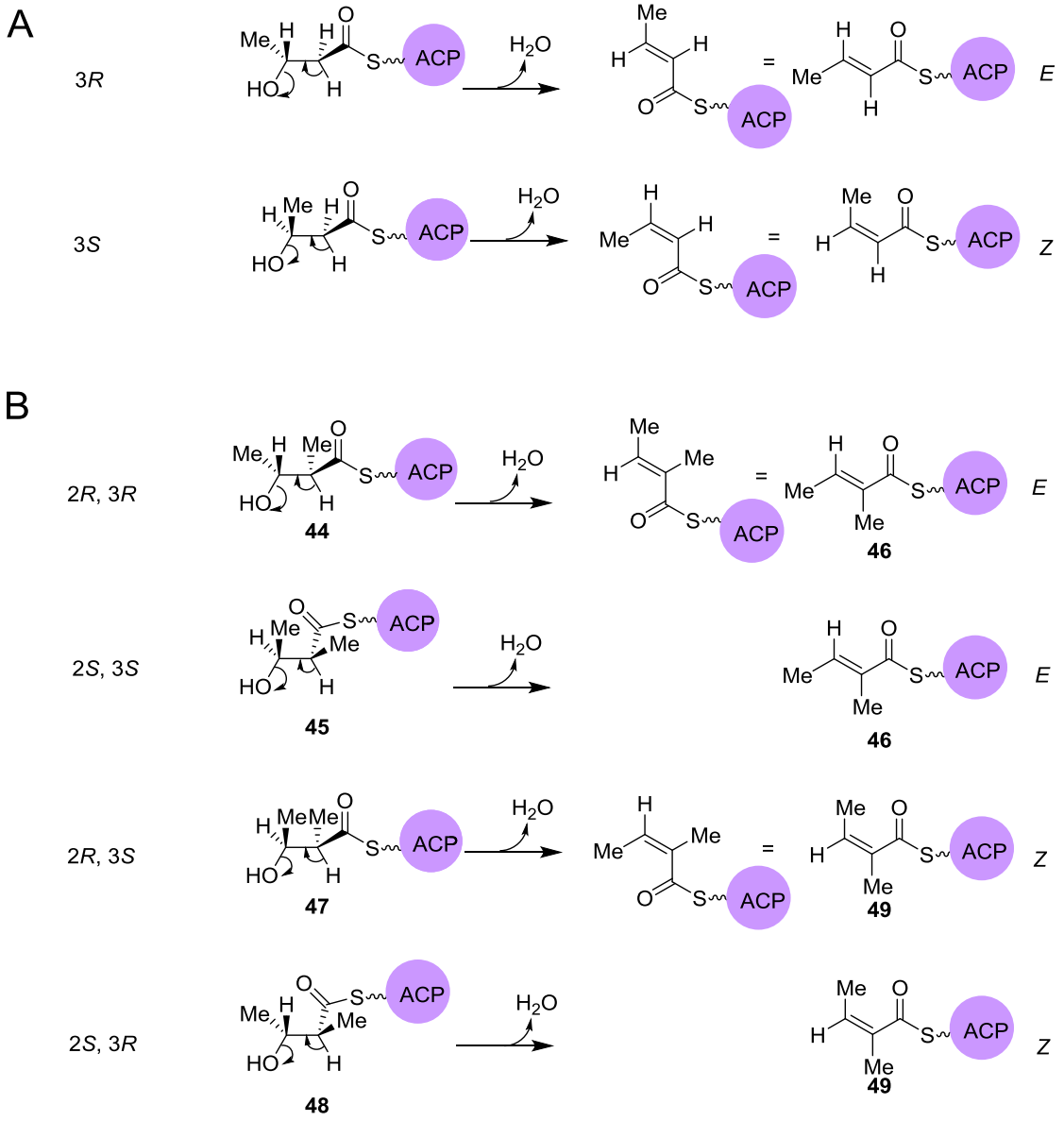
Scheme 13 – Mechanism of dehydration by DH domain.

1.5.2 Substrate specificity

There are two possible stereochemical outcomes of the dehydration of the β-hydroxy group. The double bond formed can potentially take either the *Z* or *E* geometry (Scheme 12). Currently the mechanism regarding the programming responsible for controlling the stereochemical outcome of the HR iPKS DH domain remains elusive.

It has been found that FAS DH domains catalyse the dehydration of 3*R*-β-hydroxy-acyl ACP to form an α,β-unsaturated thiolester-ACP **43** with *E* geometry.[22-24] All known DH dehydrations proceed *via* a *syn* mechanism where the C-2 proton and the C-3 hydroxyl group are lost from the same face of the molecule (Scheme 14A). In general, most polyketide compounds contain *E* double bonds,[12, 13, 17] homologous to that found in FAS,[60] although a small proportion have been

found to contain the alternative *Z* geometry, for example rifamycin (Rif10DH/RifF) and pikromycin (Pik2DH).[13, 66, 67]



Scheme 14 - **A,** Stereochemical outcomes of *syn* elimination of 3*R*-OH and 3*S*-OH of a diketide substrate in FAS; **B**, Stereochemical outcomes of *syn* elimination of 3*R*-OH and 3*S*-OH of a methylated diketide substrate in PKS. Note: mechanism is likely to be E1cb, arrows indicate syn elimination.

As enzymes have evolved over many millions of years, it is often thought that they always use the most favourable pathway when catalysing transformations.[68] In non-enzymatic elimination reactions, *syn* pathways are not favoured due to the eclipsed geometry needed for the reaction. *Anti* elimination is more favourable due to stereoelectronic effects. It is therefore not surprising that a number of hydratase-dehydratase enzymatic reactions follow *anti* elimination pathways when the abstracted proton is in the α-position. There are seven known examples including fumarate hydratase, aconitate hydratase and enolase.[68]

However, it has been found that not all enzymatic dehydration reactions follow the favourable *anti* pathway.[68] A group of dehydratases produce *syn* elimination of water, when the abstracted proton is α to a carbonyl. This group includes enoyl-coenzyme-A (CoA) hydratase, FAS and β-hydroxydecanoyl thiolester dehydratase. Work by Mohrig indicates that although *syn* elimination of water with thiolester substrates may not be the most chemically efficient pathway, it appears to be dependent on historically evolved pathways.[68] As PKS have common evolutionary routes and a high homology to FAS, it is reasonable to assume that elimination of water in PKS proceeds *via syn* elimination.

The elimination pathway taken by the PKS affects the stereochemical outcome of the product. It is generally assumed that PKS DH follows the *syn* elimination pathway. This assumption has allowed modelling of substrates in order to obtain insights into the relationship between substrate and product stereochemistry.[17] It has been shown that in α-methylated substrates chirality of the α-methyl and β-OH plays a part in determining the geometry of the double bond (Scheme 14).[17, 60] If the α-methyl and β-OH have *anti* stereochemistry, 2*R*,3*R*-**44** or 2*S*,3*S*-**45**, a *syn* elimination will result in an *E* geometry alkene **46**. *Syn* stereoisomers 2*R*,3*S*-**47** or 2*S*,3*R*-**48** results in a *Z* alkene **49**. The stereochemistry of the β-OH is dictated by KR ketoreduction.[69]

In mFAS the *C*-Met domain is present but inactive, but this is not the case in some fungal PKS. The presence of a functional *C*-Met domain adds an additional level of programming. The presence of a methylated α-position may affect the stereochemical outcome of the dehydration reaction.

Alkene geometry is not always dictated by the stereochemical outcome of the KR reduction.[60] While FAS DH domains FabA and FabZ both catalyse the dehydration of β-hydroxyacyl-ACP substrate,[70, 71] FabA is also responsible for isomerising *trans*-2-enoyl fatty acid to *cis*-3-enoyl fatty acids.[70] It has been shown that in some cases the geometry of *E* alkene arises from exogenous enzymes. One such example is the DH domain of borrelidin PKS.[60] Borrelidin **50** is synthesised by a mPKS which contains five DH domains in total and the final structure contains both a *Z* and an *E* alkene (Figure 14). The second module gives rise to an *E* alkene **51** and the third gives rise to a *Z* alkene **52**. Assays involving the isolated DH domains from the modules responsible for the formation of these double bonds (Scheme 15), module 2 and 3, found that only the 2*R*, 3*R*-**53** diastereomer was accepted and turned over by the enzyme. 2*S*, *3S*-**54**, 2*R*, *3S*-**55**, 2*S*, 3*R*-**56** diastereomers were not converted at all (Scheme 15). It was found that only the *E* product **57** was formed. These results indicate that in this case, formation of the *Z* geometry is unlikely to be due the outcome of the KR reduction alone.

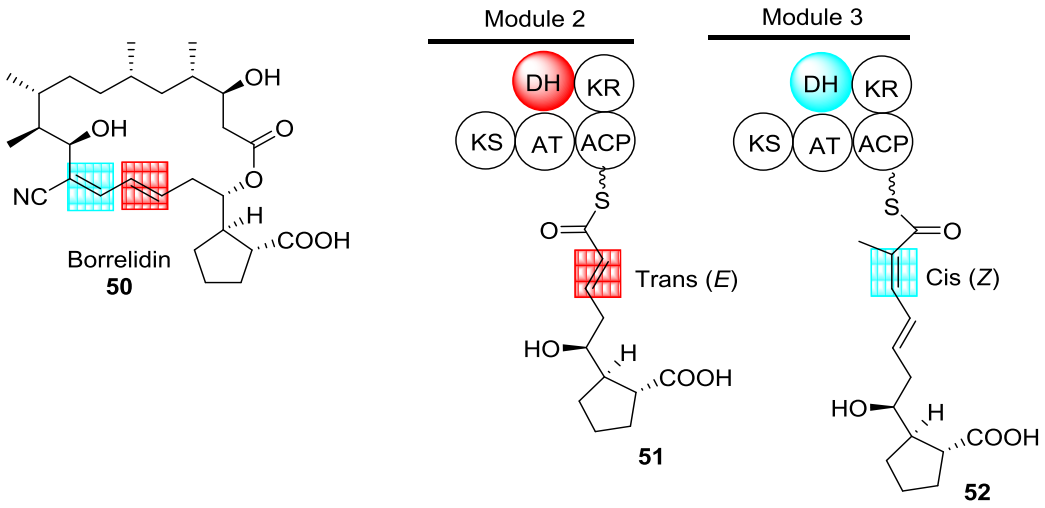
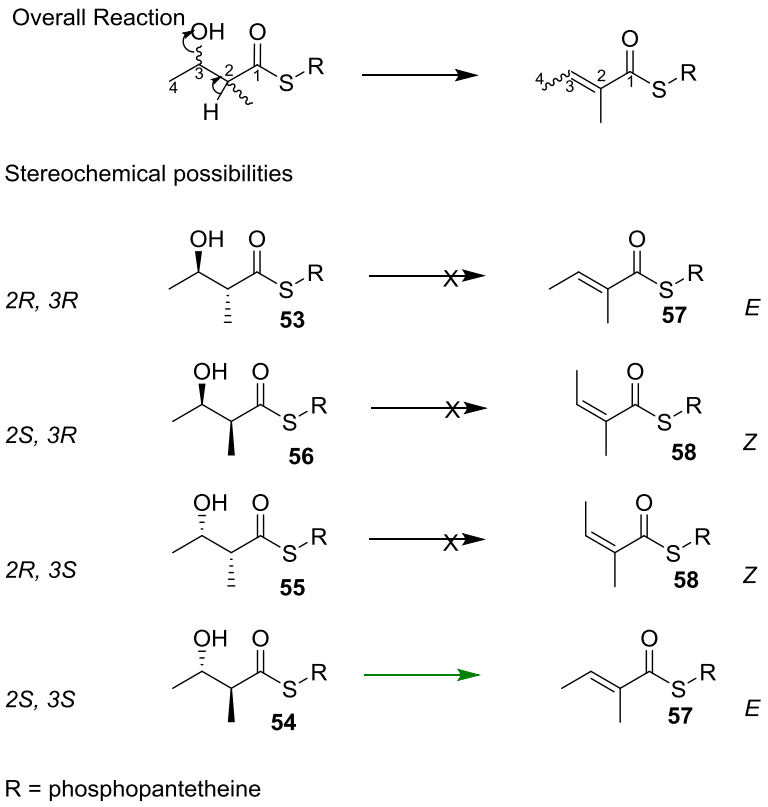


Figure 14- Borrelidin synthesis on a mPKS.[60]

The *Z* alkene could arise from a number of different routes.[60] All the double bonds could be first synthesised as *E* geometry followed by an *E* to *Z* isomerisation once the final product has been formed. The intervention of exogenous enzymes which bind to the module catalysing the formation of the *E* double bond, such as in avermectin PKS,[72, 73] cannot be ruled out. Modest alterations in enzyme architecture or substrate can affect the stereochemical outcome. If a phosphopantetheine or full ACP based substrate were to be used as a substrate instead of the *N*-acetylcysteamine (SNAC) based substrate, a different geometry double bond may be formed.



Scheme 15 - Stereochemical outcome of Borrelidin DH assay.

There is still a great deal to discover and understand about the inner workings and programming in the DH domain. Literature searches have not identified any work published investigating the programming of the iterative dehydratase domains.

1.5.3 Structures of DH Proteins

A number of FAS DH domain crystal structures have been obtained. These include the bacterial FAS DH FabA[70] and FabZ,[71] mammalian FAS DH[48] and fungal FAS DH[33, 74] domains. In addition a number of PKS DH domains have been crystallised and their structures solved. These include multiple curacin PKS DH domains, CurFDH, CurHDH, CurJDH and CurKDH.[17] The DH domain from the fourth module of erythromycin, Ery4DH has also been crystallised,[6] as has the terminal DH from rifamycin Rif10DH.[13] Most recently, the PKS DH domain from mycocerosic acid (MASDH) has been crystallised.[75] Investigation into the structure of FAS and PKS DH domains has increased the understanding of the DH active site, how the dehydration occurs and how the stereochemical outcome of the reaction is controlled.

1.5.4 Bacterial FAS DH

FabA and FabZ are standalone dehydratases which are involved in the synthesis of fatty acids in the type II FAS pathway found in bacteria.[70] They are homologous to each other and they are responsible for catalysing the dehydration of hydroxylacyl-ACP to the *trans*-2-enoyl-ACP. The protein sequences of FabA and FabZ were aligned using clustal omega (Figure 15),[76, 77]

```
bA 4FQ9      GMTKQHAFTREDLLRCSRGELFGPGN---------AQLP-APNMLMIDRIVHISDVGGKY  50
FabZ 1U1Z      ------MGSSHHHHHHS-SGLVPRGSHMMDINEIREYLPHRYPFLLVDRVVELDIEGKR-          52
                     : ..  : * . *.  *.           **    :*::**:*.::  * :

FabA_4FQ9      GKGELVAELDINPDLWFFACHFEGDPVMPGCLGLDAMWQLVGFYLGWQ---GNPGRGR--          105
FabZ_1U1Z      ----IRAYKNVSINEPFFNGHFPEHPIMPGVLIIEAMAQAAGIL-GFKMLDVKPADGTLY          107
                   : *  ::. :  **  **  .*:*** * ::** * .*:  *::    :*. *

FabA_4FQ9      -ALGSGEVKFFGQVLPTAKKVTY--NIHIKRTINRSLVLAIADGTVSVDGREIYSAEGL-          161
FabZ_1U1Z      YFVGSDKLRFRQPVLPGDQLQLHAKFISVKRSIWKF------DCHATVDDKPVCSAEIIC          161
                 :** :::*   ***  :   :   * :**:* :       *  .:** : : *** :

FabA 4FQ9      ---RVGLFTSTDSF  172
FabZ_1U1Z      AERKLGS-------  168
```

Figure 15 - Protein sequences alignment of FabA (PBD: 4FQ9) and FabZ (PDB: 1U1Z). An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties as defined by clustal omega. The catalytic amino acids. H70 and D84 in FabA and H69 and E83 in FabZ are highlighted in yellow.[76-78]

The monomeric structures of FabA and FabZ are similar, comprising an α, β hotdog fold. (Figure 16).[6] Its name arises from six stranded anti-parallel β-sheets (the bun) which curves around the hydrophobic α-helix (hotdog). The α-helix (hotdog) is shown in blue and the anti-parallel β-sheets (the bun) in pink.

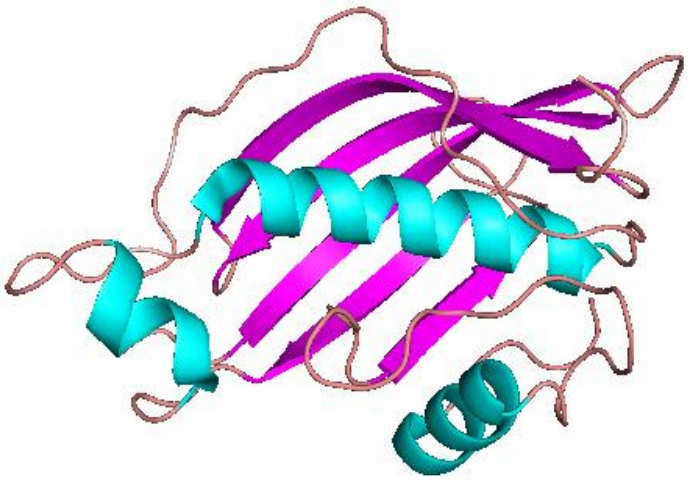
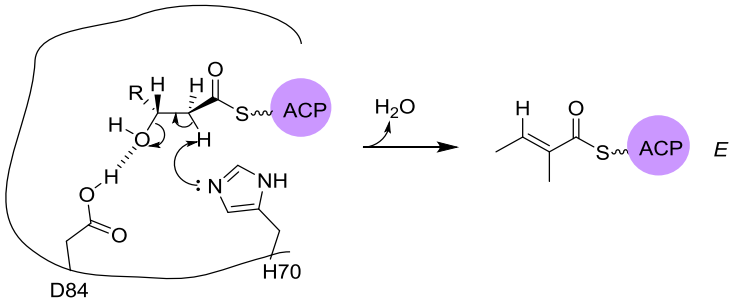


Figure 16– Crystal structure of 3-hydroxydecanoyl-acyl carrier protein dehydratase monomer (FabA) from *Pseudomonas aeruginosa*, PDB: 4FQ9.[63]

In FabA and FabZ the active site contains a histidine, (H70 in FabA and H69 in FabZ) which is responsible for abstraction of an α-position proton from substrate[70, 71] and an acid (D84 in FabA and E84 in FabZ). In FabA, D84 is part of a large hydrogen bonding network which is thought to help facilitate the protonation of the β-hydroxyl group.



Scheme 16 – Dehydration in FabA, 3*R*-hydroxyl group of the substrate mimic is able to form a hydrogen bond with the D84.

Both FabA and FabZ dehydrate the 3*R*-hydroxylacyl-ACP to the *trans*-2-enoylacyl-ACP. The crystallisation of FabA with a 3*R*-hydroxylacyl substrate mimic reported by Moynie *et al.* rationalises the stereochemical outcome of this reaction.[70] The 3*R*-hydroxyl group of the substrate mimic is positioned so hydrogen bonding with D84 can occur (Scheme 16). A 3*S*-hydroxyl group would be positioned away from D84 and hydrogen bonding would be unable to take place, resulting in the dehydration not occurring.

1.5.5 mFAS and PKS DH

Like bacterial FAS DH, mFAS and PKS DH also contain the hotdog fold structure, although in mFAS and PKS the DH consist of dimers, with each monomer containing two repeats of the hotdog fold.[17] These double hotdog folds are thought to be products of gene duplication and fusion.[17, 79]

The protein sequences of the mFAS DH and the PKS DH domains from, CurF, CurH, CurJ and CurK,[17] Ery4,[6] Rif10,[13] and MAS[75] were aligned using clustal omega (Figure 17, full alignment shown in Appendix 1),[76, 77] amino acid numbering will be based on this sequence alignment.

```
SQTKS_DH          RHFIRVQD-IPWIRDHVVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVAGYRLAEVD        88
mFAS              KFDVSPESPDHYLVDHCIDGRVLFPGTGYLWLTWKTLARALSQNL----EETPVVFEDVT       102
CurHDH_3KG7       HQDINLNN-HPWIGDHRVYDTPVIPGVSYIAMTLAAV-------------GVPAAVEDIN        72
CurJDH_3KG8       ESYFSTEN-LPFLADHIVYEQVVVPGASHISLLLAAASLTF--------AATECQIEDIL        77
CurFDH_3KG6       QSYIGAES-PGYLNHHQVFGKVLFPSTGYLEIAASAGKSLF---T----SQEQVVVSDVD        74
CurKDH_3KG9       QSYLTAES-PAYLSQHQVFNKVLFPATGYLEIAAAVGKNLL---T----TGEQVVVSDVT        79
MASDH_5BP2        QGDVGTEA-HPWLSDHRVHQVAVLPGAAYCEMALAAVTPVL---------GDTGEVHDLK        88
EryDH_3EL6        TGRLSTDE-QPWLAEHVVGGRTLVPGSVLVDLALAAGEDVG---------LP--VLEELV       102
RifDH_4LN9        TSRLSLRS-HPWLADHAVRDVVIVPGTGLVELAVRAGDEAG---------CP--VLDELV       103
                     .       :: .* :    : *.     :   .                 . ::

SQTKS_DH          ILRAMLIPDTSEGLEAHISLRPCSTKLLLTNEWYDFCVSSVGDD---------------D       133
mFAS              LHQATILPKTGTV-SLEVRLLEA--------SH-AFEVSD-------------------S       133
CurHDH_3KG7       FQQPLFLAESNTTRETQLMLHTA-----DNVGK-QF------------VEVFSRDGAKQE       114
CurJDH_3KG8       FPQALAIPEQGVR-TVQVVLT-P-----QNNSF-SFQVISFDDSLESQINQVSNNGSHIS       129
CurFDH_3KG6       ILQSLVIPETEIK-TVQTVVSFA-----ENNSY-KFEIFSPSEG----------ENQQTP       117
CurKDH_3KG9       IVRGLVIPETDIK-TVQTVISTL-----ENNSY-KLEIFSTSEG----------DNQQAN       122
MASDH_5BP2        FHDMLLLDDATPV-WVSAAVTAP-----GTAEF-GVETHQ-----------------SG       123
EryDH_3EL6        LQRPLVLAGAG-A-LLRMSVGAP-----DESGRRTIDVHAAEDV----------ADLADA       145
RifDH_4LN9        IEAPLVVPRRGGV-RVQVALGGP-----ADDGSRTVDVFSLRED--------------AD       143
                  :     :           :              .

SQTKS_DH          KFVDHCRGRITIEFDTSGSADTPRTSLRERSRSTGLMRSVDPSNLYSFLRAQGIYHGPIF       193
mFAS              NGSLIASGKVYQWESPDPKLFDTRAAVD--PADSTAEFRLSQGDVYKDLRLRGYDYGPFF       191
CurHDH_3KG7       EWQQHASMSVSENPPPPPTLSVD---I---PALCEQLRPLDTDTLTEIYASISLVYGPML       168
CurJDH_3KG8       DWAVHATGKLSVAN--AEQSLIPLEEI---QARCSQ--KIDSAEIYQHLWDRQIHLGQSF       182
CurFDH_3KG6       QWVLHAQGKIYTEPTRNSQAKIDLEKY---QAECSQ--AIEIEEHYREYRSKGIDYGSSF       172
CurKDH_3KG9       QWTLHAEGKIFLDSTTNTKAKIDLEQY---QRECSQ--VIDIQQHYQQFKSRGIDYGNSF       177
MASDH_5BP2        DRTQRATAVLRGDVDAERPAAHSIDAL---LAAHPN--RVDGDELRAGFGTVGIGHGAAF       178
EryDH_3EL6        QWSQHATGTLAQGV---AAGPRDTEQW---PPEDAV--RIPLDDHYDGLAEQGYEYGPSF       197
RifDH_4LN9        SWLRHATGVLVPENRPRGTAAFDFAAW---PPPEAK--PVDLTGAYDVLADVGYGYGPTF       198
                  .    .   :                              :                * :

SQTKS_DH          QNLKTISSRKDHSES----SFVVANTASVMPNGFQSPHVIHPTTLDSIFQGAYTA-LPGA       248
mFAS              QLVLESDLEGNRG------R--------------LQWNDSWVSFLDAMLHMSILAP----       227
CurHDH_3KG7       QAVRQAWIGEETS----LLEIEVPKALA----FQLAGEPIHPVLIDACTRLT-PDLFDFS       219
CurJDH_3KG8       RWIEQVWLGEGEV----LCQMKVPKTIL----N-TTKYQLHPTLVDSCFQSIIALVLDQS       233
CurFDH_3KG6       QGIKQLWKGQGKA----LGEMAFPEELT----AQLADYQLHPALLDAAFQIV-SYAIPHT       223
CurKDH_3KG9       QGIKQLWKGQGKA----LGKIALPEEIA----GQATDYQLHPALLDAALQIL-GHAIGNT       228
MASDH_5BP2        AGLSEAYVATAAE-PTVVAAVALPGPLR----SGQRGYTVHPALLDACFQSVIAHPEV--       231
EryDH_3EL6        QALRAAWRKDDSV----YAEVSIA--------ADEEGYAFHPVLLDAVAQTLSLGALG--       243
RifDH_4LN9        RAVRAVWRRGSGNTTETFAEIALPEDAR----AEAGRFGIHPALLDAALHSTMVSAAADT       254
                    :                                          :*:  :
```

Figure 17 - Protein sequences alignment of SQTKS DH, mFAS (PDB: 2VZ9), CurH (PDB: 3KG7), CurJ (PDB: 3KG8), CurF (PDB: 3KG6), CurK (PDB: 3KG9), MASDH (PDB: 5BP2), Ery4DH (PDB: 3EL6) and RifDH (PDB: 4LN9). An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties. A . (period) indicates conservation between groups of weakly similar properties as defined by clustal omega. The catalytic amino acids. H70 and D84 in FabA and H69 and E84 in FabZ are highlighted in yellow.[76-78]

As discussed in Section 1.3.1, mFAS forms an "X" shaped dimer with the dimers interfacing down the middle of the X, resulting in the main interfacing interactions occurring at the KS, ER and DH. The DH domain dimerises through homophilic interactions between the hotdog folds. Each DH domain monomer in mFAS adopts a pseudo dimeric double hot dog fold,[48] which contains a single composite active site (Figure 18, A and B).[48] The active site is formed by H62 on the N-terminal hot dog fold and D217 on the C-terminal hot dog fold (mFAS numbering). H62 and D217 are the key catalytic amino acids which participate in the protonation and deprotonation of substrate.

In chicken and pig FAS, a histidine, H220 is found close to D217. In all other sequenced FAS, glutamine is found in this position. The H220 is positioned within hydrogen bonding distance of D217 and is thought to provide a stabilising effect for D217. Mutagenesis experiments showed

that the histidine and glutamine are equivalent at this position. The hydrophobic substrate binding tunnel opening is situated at the pseudo-dimer interface. The tunnel stretches through the C-terminal hot dog fold (Figure 18, C and D).

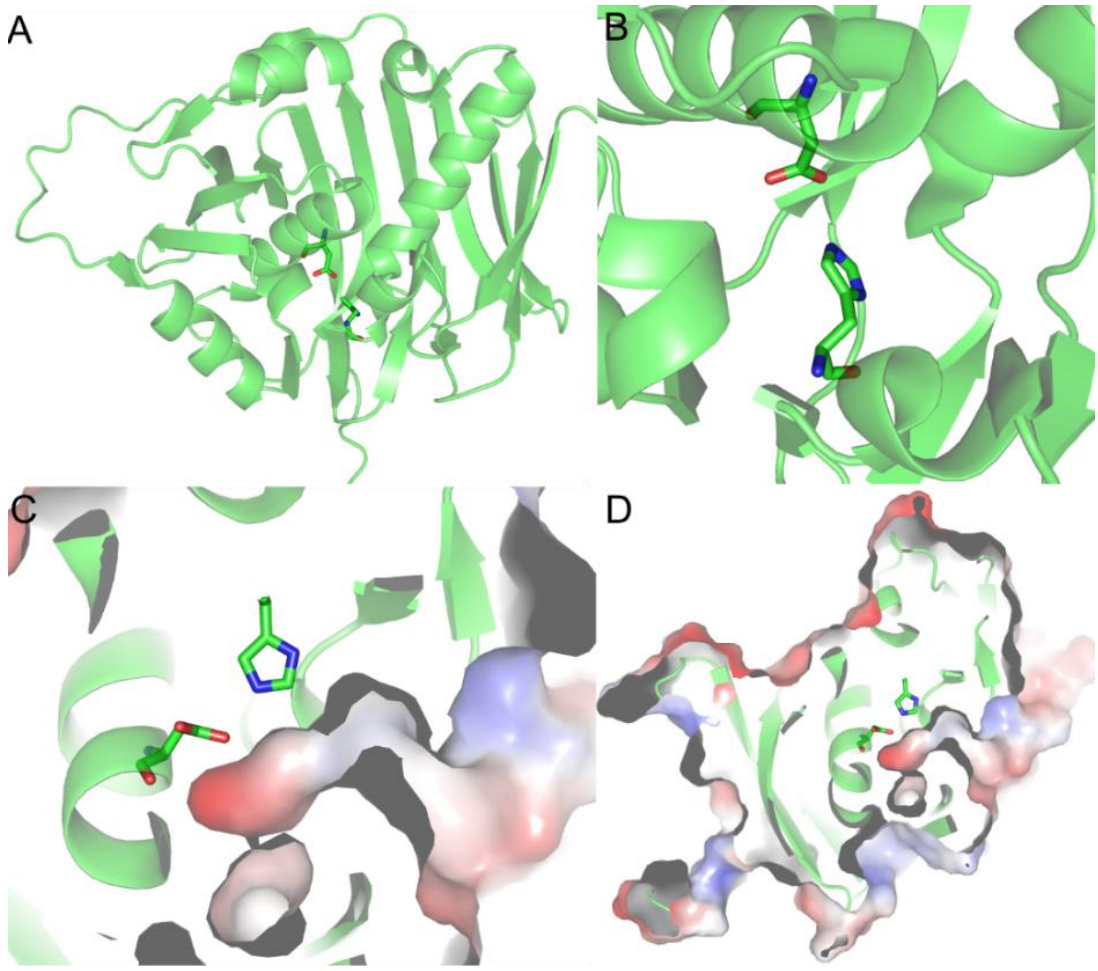
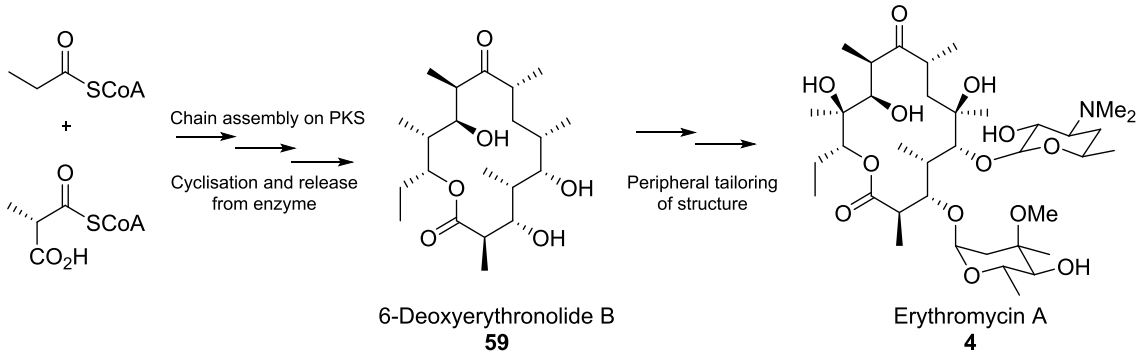


Figure 18: **A**, Overall crystal structure of mFAS DH domain monomer; **B**, catalytic amino acids H62 and D217 in active site; **C**, Side view of mFAS substrate tunnel with H62 and C217 shown; **D**, Cut through of overall crystal structure of mFAS DH domain monomer showing substrate binding tunnel and catalytic amino acids. (PDB: 2VZ8).[48]

1.5.6 Erythromycin, Ery4DH

Erythromycin **4** is a macrolide antibiotic, the biosynthesis of which can be divided into two parts. A mPKS produces 6-deoxyerythronilide B **59** by catalysing the sequential condensation and tailoring of one propionyl CoA starter unit and six methylmalonyl CoA extender units. Once 6-deoxyerythronoilide B has been produced and released from the PKS, a series of further tailoring reactions decorate the peripheral structure (Scheme 17).[6, 12, 40, 55]



Scheme 17 – Biosynthesis of erythromycin.

The PKS consists of three large polypeptides, DEBS1, DEBS2 and DEBS3. Each polypeptide contains two PKS modules (Scheme 18).[21] Of the 6 modules, only module 4 contains a DH domain, known as Ery4DH. This catalytic domain has been isolated as a stand-alone protein by Keating-Clay[6] and its structure solved.

Ery4DH is formed of a single polypeptide chain which folds to form a double hotdog fold and contains the HXXXGXXXXP signature sequence motif found in all PKS DH domains near the N-terminus. Within the active site there are a number of key amino acids, including the catalytic histidine (H69) required for activity. H69 is highly conserved and invariant. Nearby amino acids (Figure 19), which are also highly conserved, contribute to a hydrogen bonding network which ensures the correct orientation of the histidine. A turn in the protein structure allows P78 to make contact with H69 *via* Van der Waals bonding interactions. An additional amino acid, L76, hydrogen bonds to H69 helping to maintain the correct orientation for catalytic efficiency.

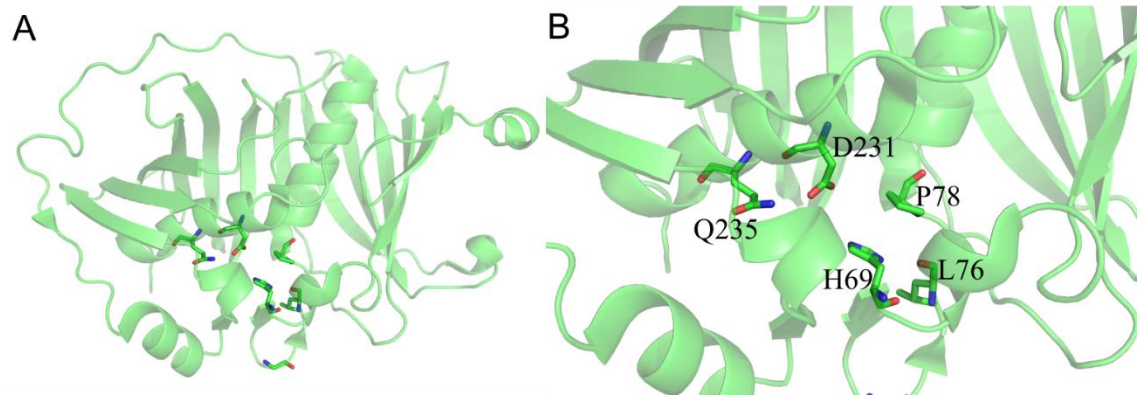
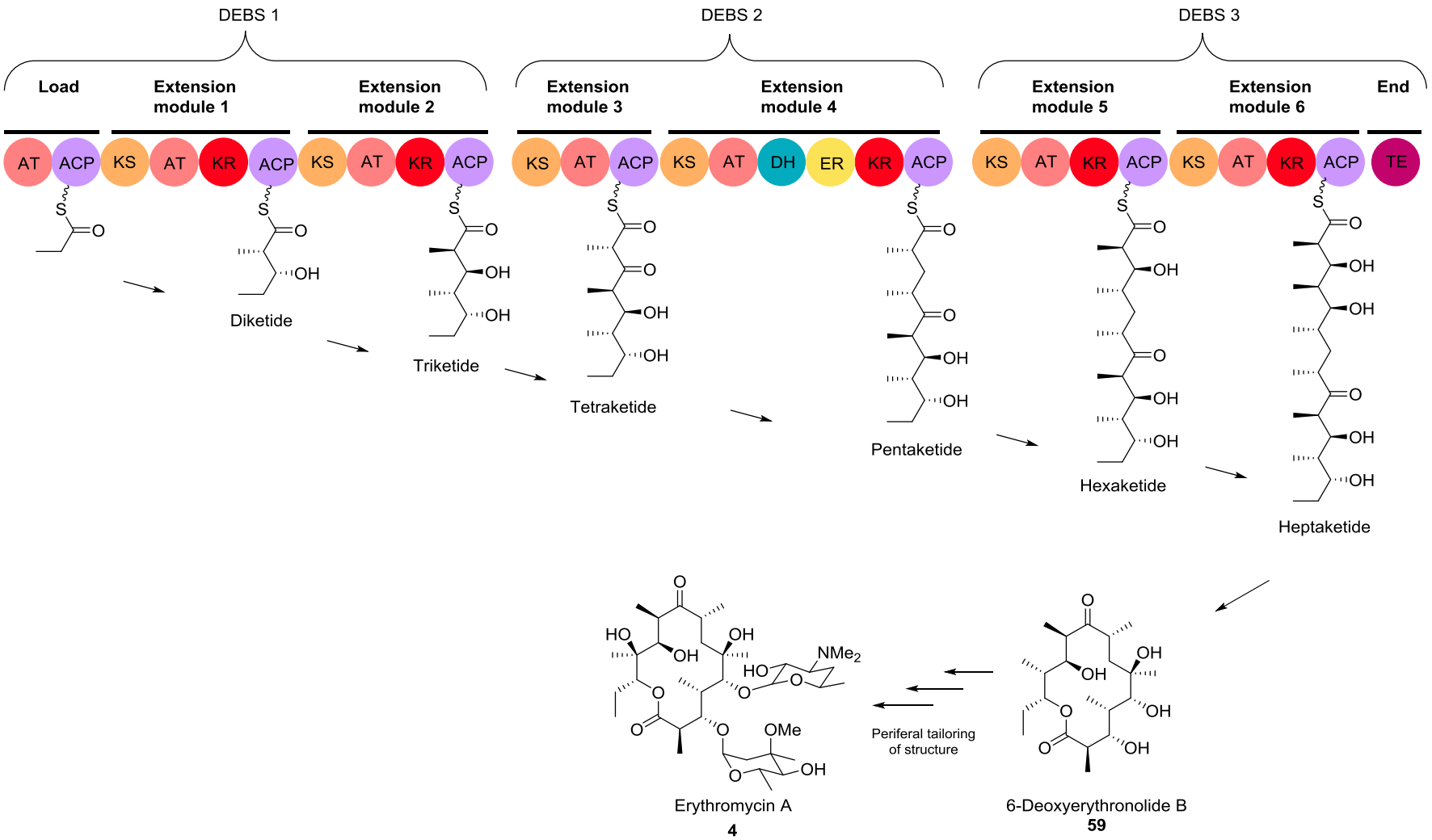
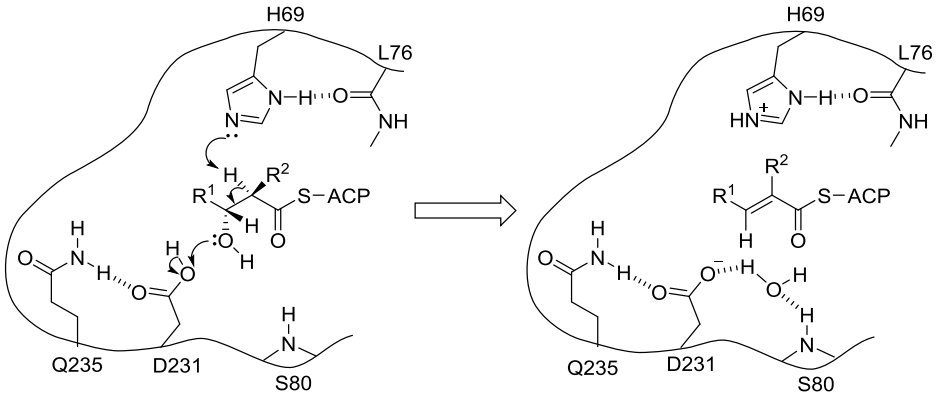


Figure 19 – **A.** Overall crystal structure of Ery4DH **B.** Key catalytic amino acids H69, D231. Supporting amino acids G73, P78, L76, and Q235.



Scheme 18 – Erythromycin PKS DEBS, three large polypeptides, DEBS1, DEBS2 and DEBS3, each polypeptide contains two PKS modules 30

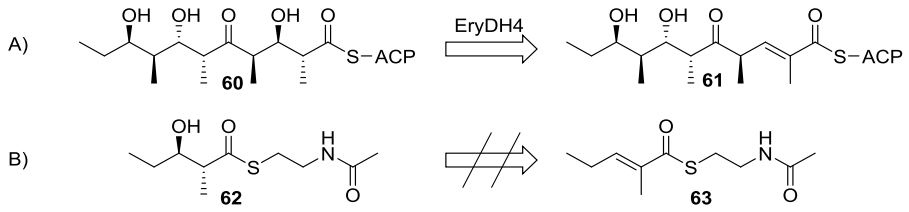
The second catalytic amino acid, aspartic acid is highly conserved in type I PKS DH domains, whereas in type II dehydratases this role is taken by glutamic acid (Scheme 19). Within the active site of Ery4DH, D231 is anchored in position by hydrogen bonding to Q235. In the crystal structure, water was found bound to S80. It was hypothesised that this water was the result of dehydration of the substrate's acyl chain.



R^1 = Growing PKS chain
R^2 = H or Me

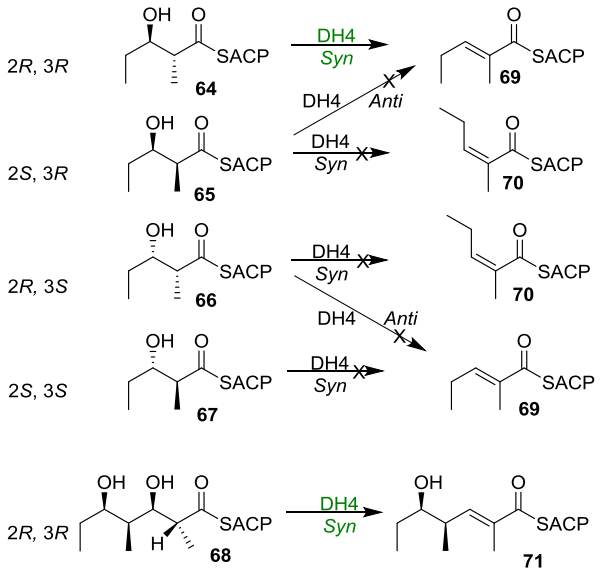
Scheme 19 – Cartoon of substrate docked into the active site and the interaction of amino acids in Ery4DH active site.

Ery4DH's natural substrate, the pentaketide **60,** is dehydrated by the DH domain to form **61** (Scheme 20, A). Substrate **60** is difficult to isolate so assays carried out to assess the activity of Ery4DH utilised a simplified substrate, the 2*R*,3*R*-3-diketide SNAC **62**.[6] However, analysis by liquid chromatography-mass spectrometry (LC-MS) did not observe formation of product **63** (Scheme 20, B).



Scheme 20 – A. Natural substrate **60** for Ery4DH which is dehydrated by Ery4DH; B.[6] Substrate mimic **62**, which did not undergo dehydration.

A later paper by Cane *et al.*[12] repeated the assays of Ery4DH using ACP bound substrate mimics **64-68** (Scheme 21). Incubation of the recombinant Ery4DH with 2*R*,3*R*-**64** resulted in the formation of dehydrated product **69**. A reverse catalytic reaction was observed when Ery4DH was incubated with **69**, resulting in a ~3:1 equilibrium of **64** and **69**. The results of these assays highlight the importance of the presence of ACP DH activity.



Scheme 21 - Cane et al. assays to assess the selectivity of Ery4DH and the outcome.[12]

Further assays were carried out by Cane *et al.* to assess the selectivity of Ery4DH.[12] The enzyme was found to be highly selective. Only the 2*R*,3*R*-ACP **64** was dehydrated to give the *E*-**69**-ACP product. The 2*S*,3*R*-**65**-ACP, 2*R*,3*S*-**66**-ACP and 2*S*,3*S*-**67**-ACP substrates did not undergo dehydration to give the *E*-**69**-ACP or *Z*-**70**-ACP products. 2*R*,3*R*,4*S*,5*R*-2,4-dimethyl-3,5-dihydroxyheptanoyl-ACP **68**, was also incubated with Ery4DH to further confirm the stereoselectivity of the DH domain, and this substrate underwent dehydration to form product **71**.

1.5.7 Rifamycin Rif10DH

Rifamycin **72**,[80, 81] is an ansamycin antibiotic and a number of semi synthetic derivatives of rifamycin have been clinically approved as an effective treatment for tuberculosis. These include rifapentine, rifampicin and rifabutin.[81] Rifamycin is produced by the actinomycete *Amycolatopsis mediterranei* and contains a number of unique structural features, including an amino naphthoquinone, a macrolactam ring and a tri-substituted *Z* alkene.

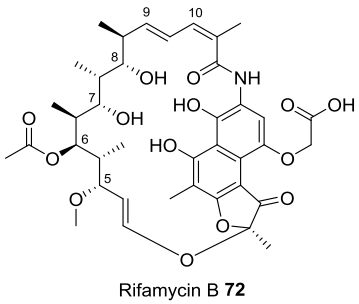


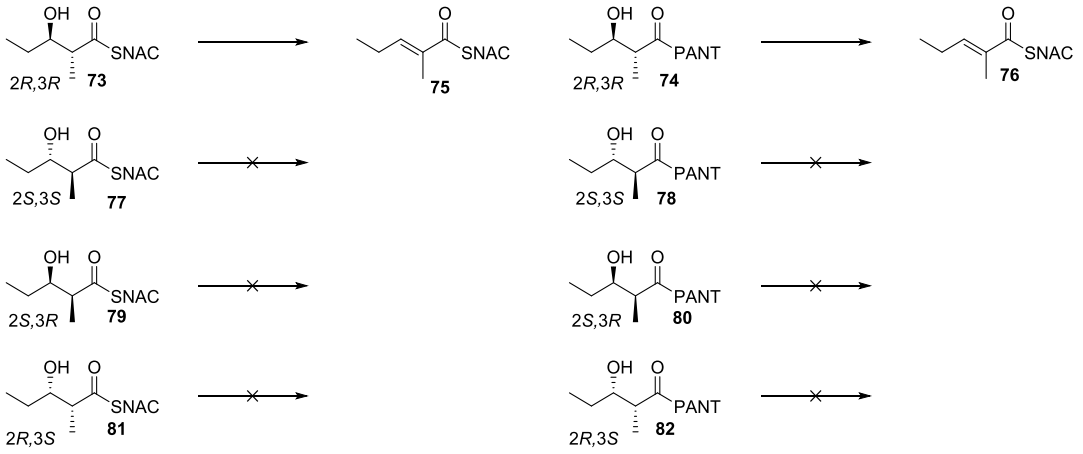
Figure 20 – Position numbering refers to the activity by Rif modules.

Rifamycin **72** is produced by a type I mPKS. The 10 PKS modules extend the 3-amino-5-hydroxybenzoic acid starter unit with two malonate and eight methyl malonate extender units to

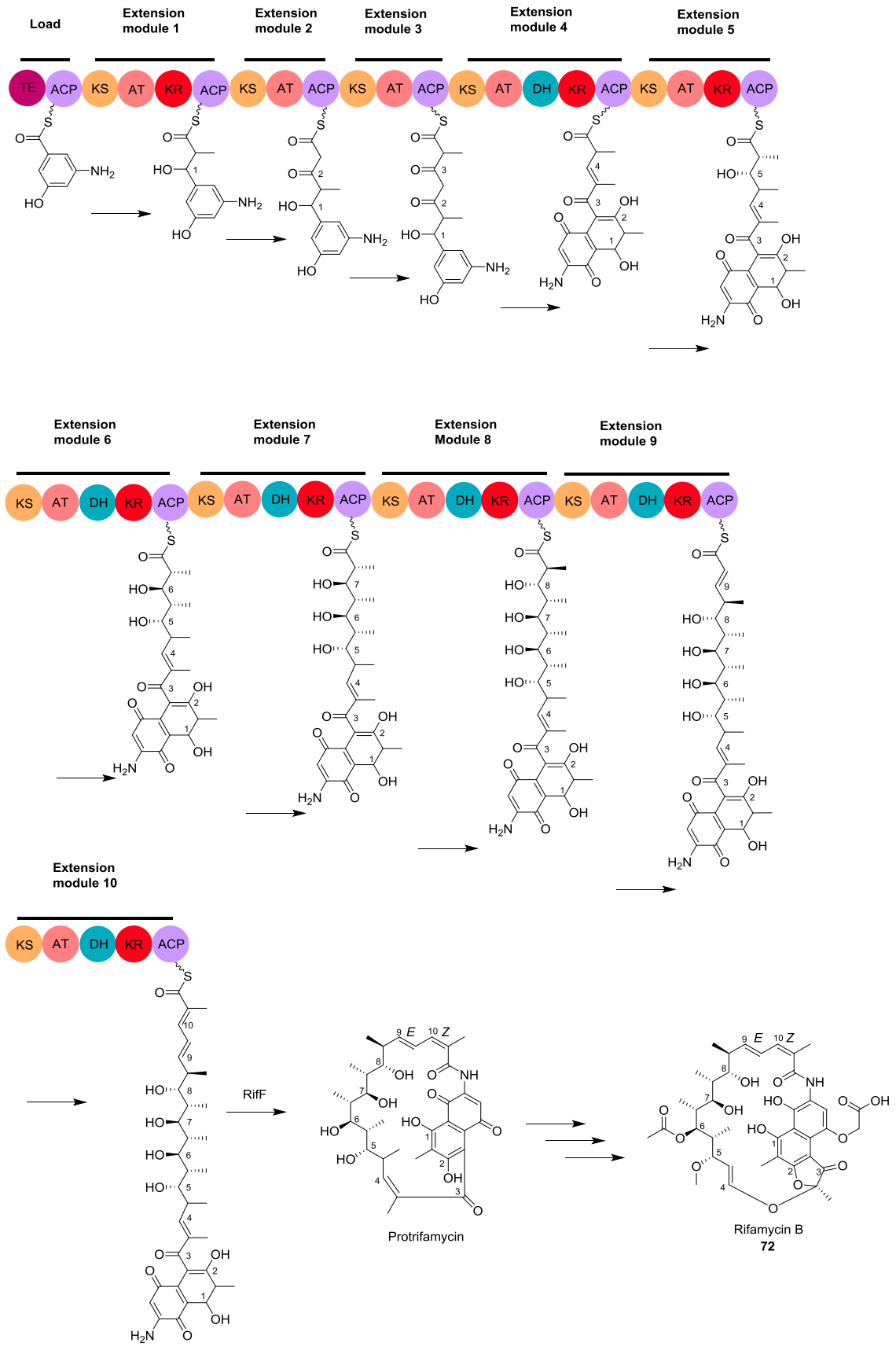
form the rifamycin skeleton (Scheme 23). Eight of the ten modules contain putative DH domains with 41-49% identity to Ery4DH. Of these eight, six possess a region which shows good similarity to a DH active site (modules 4, 6, 7, 8, 9 and 10). The presence of DH domains, which could potentially be active in Rif modules 6, 7 and 8 was unexpected as the corresponding positions in rifamycin B **72** are hydroxylated (Figure 20). It is possible these DH domains are inactive or, if they are active the hydroxyl groups are added at a later time. Rifamycin PKS does not contain a thiolesterase (TE) domain which would be responsible for the release of the fully processed undecaketide. Instead the release is catalysed by RifF, a separately encoded amide synthase. Mutation or deletion of RifF results in the release of a 2-methyl-2-enoyl undecaketide. Two research teams investigated the effect of mutating or deleting RifF on the geometry of the double bond in the fully processed undecaketide and reported conflicting results.[82-84] It was unknown if the *Z* geometry of the undecaketide double bond is installed by Rif10DH or by RifF.

Work carried out by Gay *et al.* focussed on Rif10DH in more detail.[13] Structural and sequence alignment of the Rif10DH did not reveal any indication of whether the alkene it produced would be *Z* or *E* in the final product. Rif10DH, was then isolated as a standalone protein and crystallised.

Gay *et al*. reported that they had undertaken incubation of Rif10DH with acyl-SNAC, acyl-Pantetheine and acyl-ACP substrate mimics.[13] The exhibited selectivity towards a certain substrate varied depending on the substrate mimic's design. When Rif10DH was incubated in the presence of an acyl-SNAC or an acyl-pantetheine, the DH exhibited high selectivity towards the 2*R*,3*R*-2-methyl-3-hydroxypentanoyl isomers **73** and **74** (Scheme 22). The resulting dehydration gave the *E*-2-methylpentanoyl products **75** and **76**. The other substrate 2*S*,3*S*-**77** and **78**, 2*S*,3*R*-**79** and **80** and 2*R*,3*S*-**81** and **82** were not turned over to any measurable extent (Scheme 22).



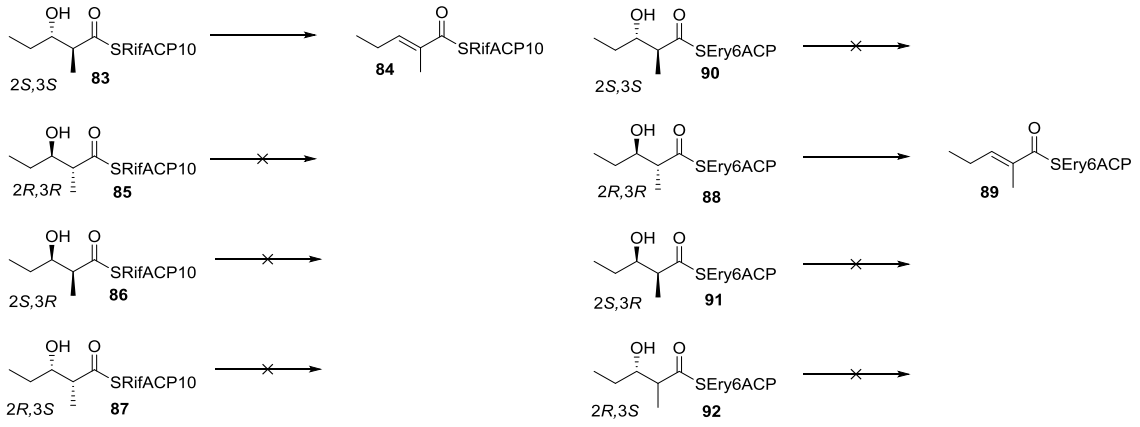
Scheme 22 – Assays carried out by Gay *et al*. investigating the selectivity of Rif10DH using acyl-SNAC, acyl-pantetheine substrate mimics.[13]



Scheme 23 – Biosynthetic pathway of rifamycin **72**, numbered positions refers to processing by Rif modules.[81]

Interestingly, when the assays were carried out using RifACP10 based mimics, a different selectivity preference was observed. Rif10DH was incubated with ACP based mimics for 1 hour. The products were then hydrolytically released using the thiolesterase, PicTE, and subsequently derivatized using TMS-diazomethane. Samples were analysed by chiral GC-MS and compared to synthetic standards as appropriate. Only the Rif10DH domain catalysed the dehydration of 2*S*,3*S*-2-methyl-3-hydroxypentanoyl-Rif10ACP **83** to form the *E*-2-methylpentanoyl-Rif10ACP **84** (Scheme 24). All other stereoisomers, 2*R*,3*R*-**85**, 2*S*,3*R*-**86**, 2*R*,3*S*-**87** were not substrates. There have been very few cases where this selectivity has been observed. The reverse reaction, *syn* hydration, was also observed when *E*-2-methyl-2-pentenoyl-Rif10ACP **84** was incubated with Rif10DH. All other characterised DH domains, such as mFAS (1.2.1) and Ery4DH (1.5.6) have catalysed only the *syn* dehydration of 2*R*,3*R*-2-methyl-3-hydroxyacyl substrates to give *E*-2-methyl-2-enoylacyl products.

On replacement of the native Rif10ACP with the ACP from erythromycin, Ery6ACP, rather than the selectivity being degraded so a number of substrates were turned over, the selectivity of Rif10DH was altered completely. 2*R*,3*R*-2-methyl-3-hydroxypentanoyl-Ery6ACP **88** became the preferred substrate and was dehydrated to form the *E*-2-methylpentanoyl-Ery6ACP **89**. Substrates **90** to **92** were not turned over to form either the *Z* or the *E* product. When the sequences of the two ACP were compared, Ery6ACP was found to be only 46% identical to Rif10ACP.



Scheme 24 - Assays carried out by Gay *et al.* investigating the selectivity of Rif10DH using acyl-EryACP6 or acyl-RifACP10 substrate mimics.[13]

The substrates used in this study were diketide-based, and therefore much shorter in length than the natural substrate of Rif10DH, a 2*S*,3*S*-2-methyl-3-hydroxyacyl-undecaketide. The authors carried out a study to assess the effect of chain length and substitution pattern on the stereoselectivity of Rif10DH. Neither chain length or substitution pattern was found to influence the exhibited stereoselectivity. It was proposed that when SNAC, pantetheine and EryACP6

based substrates dock into the active site of Rif10DH, the critical interaction between the DH and the ACP are not formed. The loss of these interactions results in the different placement of the substrate within the active site, leading to alternative substrate stereochemistry being recognised and dehydration being catalysed.

Structurally Rif10DH is highly homologous to other crystallised DH domains, such as Ery4DH and the Curacin DH domains from CurF, CurK, CurJ and CurH. Rif10DH possesses the characteristic double hotdog fold and catalytic dyad consisting of H70 and D240. In addition to the catalytic dyad Rif10DH also possesses a hydrogen bonding network which helps increase the pK_a of D240, so it can act as a general acid for the dehydration reaction. In Rif10DH H244 hydrogen bonds to D240, bridging the gap between the catalytic aspartic acid and the rest of the hydrogen bonding network. Five of the six known DH structures possess a similar hydrogen bonding network, utilising either a histidine or a glutamic acid residue. CurH is the only exception where an arginine residue is used.

1.5.8 Curacin

The anticancer lead agent curacin A **93** is produced by a type I mPKS. Curacin A **93** contains a diverse range of structural motifs including a cyclopropyl ring, a *cis* alkene, a thiazolidine ring and a terminal alkene. In total the PKS responsible for the synthesis of curacin A utilises at least 41 catalytic domains across 13 polypeptide chains. The cyclopropyl ring is installed by a hydroxymethyl glutamate (HMG)-ACP synthase. The cysteine derived thiazoline ring is installed by a NRPS-PKS in the CurF module, while the terminal alkene is installed by a sulfotransferase thiolester domain in the CurM module. The other PKS modules CurG-L are typical monomodule PKS, possessing a range of β-processing domains, with CurJ and CurL also containing *C*-methyltransferase domains.

Akey *et al.* undertook an in-depth study of the curacin DH domains, with the aim of identifying how the substrate and products are loaded and unloaded.[17] They also wished to identify the mechanism behind the formation of *cis* and *trans* alkene. Based on the mature structure of curacin A and the gene cluster, a biosynthetic pathway has been proposed by Chang *et al*.[85] It is expected that dehydration reactions occur, with DH domains present in CurG, CurH, CurI, Cur J and CurK (Figure 21). Sequence analysis carried out by Akey *et al.*, using newly available information on DH domains gained from the study of mFAS and ER, resulted in the reassignment of DH domains within the Cur PKS.[17]

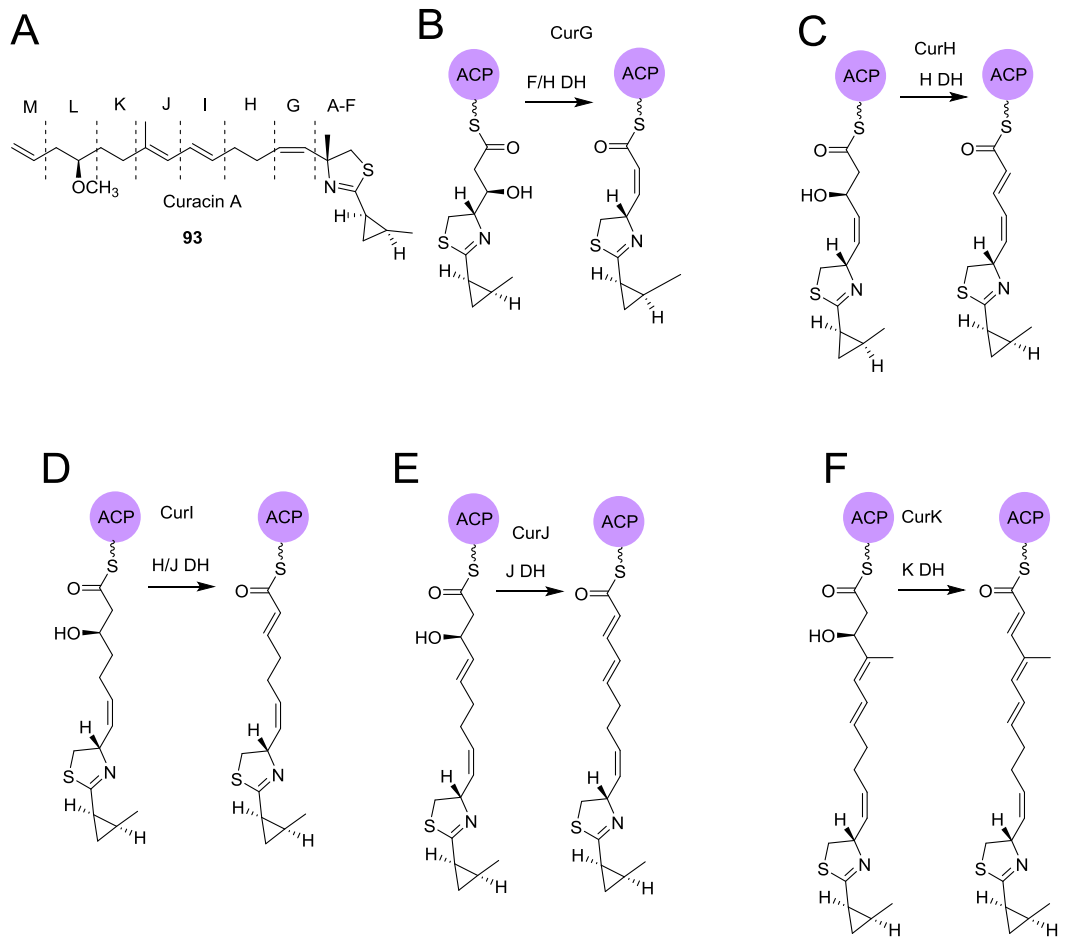


Figure 21 – A. curacin A, the single letter codes refer to the module likely to be responsible for extension and modification of the functional group. B-F. Proposed dehydration reactions catalysed occurring on intermediates of curacin A. The proposed reactions have been based on the mature structure of curacin A.

DH domains were identified in CurH, CurJ and CurK. The expected DH in CurI and CurG were not present, instead a DH was identified in CurF. The four identified DH domains were expressed as isolated soluble proteins and crystallised. The four CurDH domains had moderate levels of sequence identity to each other (21-28%), although CurF and CurK were 68% identical. Despite the low sequence identity, the four domains possess strong structural similarities. All four DH domains are dimeric, with the monomer possessing a double hotdog fold, as observed in mFAS and other crystallised PKS DH domains.[6, 17, 48]

As with Rif10DH, ERDH4 and mFAS, the active site comprises two catalytic amino acids, a histidine and aspartic acid residue (CurF: H37, D210; CurH: H41, D206; CurJ: H41, D219; CurK: H42, D215). The active site arrangement, substrate tunnel and tunnel entrance are all comparable to those found in mFAS, Ery4DH, FabA and FabZ.

The modelling of substrate docking undertaken by Akey *et al*. found that, as with other DH substrates, the histidine abstracts a proton from the α-position, while the aspartic acid residue is involved in the protonation of the β-hydroxyl group and subsequent loss of water.[17] They also identified that two water molecules present in the active site play a part in positioning of the

substrate. Comparison of all four CurDH sequences with the sequences of Ery4DH indicates the presence of the hydrogen bonding network as seen in Ery4DH (Figure 22). The proline (P78) and leucine (L76) residues observed in Ery4DH are also observed in all four CurDH domains. While the proline (CurF: P46, CurH, P50, CurJ: P50, CurK: P51) is highly conserved, the leucine is less well conserved and in CurH and CurJ a valine is found in this position (CurF: L44, CurH: V48, CurJ: V48, CurK L49, Figure 22). The proline and leucine/valine residues help ensure the correct positioning of the catalytic histidine. As found in Ery4DH the glutamine residue, which is involved in ensuring the correct orientation of the catalytic aspartic acid, is also conserved in the CurD domains with the exception of CurHDH, which possess an arginine residue at this position. (CurF: Q214, CurH: R210, CurJ: Q223, CurK: Q219, Figure 22).

The modelling also indicated that while the positioning of the substrate α-protons and the β-hydroxyl group is key for the dehydration reaction, exquisite substrate recognition is unnecessary due to the substrate being delivered to the active site by the ACP. As a consequence, the substrate tunnel lacks hydrogen bonding groups and is largely hydrophobic and lacking shapes tailored to the structure of the substrate.

```
SQTKS_DH        RHFIRVQD-IPWIRDHVVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVAGYRLAEVD       88
mFAS            KFDVSPESPDHYLVDHCIDGRVLFPGTGYLWLTWKTLARALSQNL----EETPVVFEDVT       102
CurHDH_3KG7     HQDINLNN-HPWIGDHRVYDTPVIPGVSYIAMTLAAV-------------GVPAAVEDIN       72
CurJDH_3KG8     ESYFSTEN-LPFLADHIVYEQVVVPGASHISLLLAAASLTF--------AATECQIEDIL       77
CurFDH_3KG6     QSYIGAES-PGYLNHHQVFGKVLFPSTGYLEIAASAGKSLF---T----SQEQVVVSDVD       74
CurKDH_3KG9     QSYLTAES-PAYLSQHQVFNKVLFPATGYLEIAAAVGKNLL---T----TGEQVVVSDVT       79
MASDH_5BP2      QGDVGTEA-HPWLSDHRVHQVAVLPGAAYCEMALAAVTPVL----------GDTGEVHDLK       88
EryDH_3EL6      TGRLSTDE-QPWLAEHVVGGRTLVPGSVLVDLALAAGEDVG----------LP--VLEELV       102
RifDH_4LN9      TSRLSLRS-HPWLADHAVRDVVIVPGTGLVELAVRAGDEAG----------CP--VLDELV       103
                   .       :: .* :    : *.     :   .                  . ::

SQTKS_DH        QNLKTISSRKDHSES----SFVVANTASVMPNGFQSPHVIHPTTLDSIFQGAYTA-LPGA       248
mFAS            QLVLESDLEGNRG------R--------------LQWNDSWVSFLDAMLHMSILAP----       227
CurHDH_3KG7     QAVRQAWIGEETS----LLEIEVPKALA----FQLAGEPIHPVLIDACTRLT-PDLFDFS       219
CurJDH_3KG8     RWIEQVWLGEGEV----LCQMKVPKTIL----N-TTKYQLHPTLVDSCFQSIIALVLDQS       233
CurFDH_3KG6     QGIKQLWKGQGKA----LGEMAFPEELT----AQLADYQLHPALLDAAFQIV-SYAIPHT       223
CurKDH_3KG9     QGIKQLWKGQGKA----LGKIALPEEIA----GQATDYQLHPALLDAALQIL-GHAIGNT       228
MASDH_5BP2      AGLSEAYVATAAE-PTVVAAVALPGPLR----SGQRGYTVHPALLDACFQSVIAHPEV--       231
EryDH_3EL6      QALRAAWRKDDSV----YAEVSIA--------ADEEGYAFHPVLLDAVAQTLSLGALG--       243
RifDH_4LN9      RAVRAVWRRGSGNTTETFAEIALPEDAR----AEAGRFGIHPALLDAALHSTMVSAAADT       254
                  :                                        :*:  :
```

Figure 22 – Sequence comparison of crystallised DH domains, conserved catalytic amino acids histidine and aspartic acid highlighted in yellow. Conserved amino acids involved in the hydrogen bonding network are highlighted in blue.

1.5.9 Mycocerosic acid synthase[75]

Mycocerosic acid (MAS) is a fully reducing iPKS found in *mycobacterium smegmatis*. MAS utilises C_{12} to C_{20} fatty acids as starter units.[75] The start unit is iteratively elongated one to four times using methyl-malonyl-CoA extender units to form C_{20} to C_{28} branched fatty acids. These lipids are vital for the formation of the mycobacterial cell envelope. As with other PKS, MAS is comprised of two regions, a condensing region (ACP, KS, AT), and an extension region (DH, ψKR, ER and ER). MAS iteratively does not contain a CMeT domain, the branched methyl groups arise from the incorporation of methylated extender units. Work by Maier *et al.* crystallised MAS modifying and condensing region separately. [75] The MAS DH is the most

recently crystallised DH domain available, it is also the only iterative PKS DH domain to be crystallised at this time (Figure 23). MAS DH possess many features which have been observed in the other crystallised DH domains. The MAS DH connects the modifying and condensing regions together with flexible linkers. As seen in FAS, the modifying region dimerizes along an interphase formed by the DH and ER domains. The DH dimer in MAS lacks the V-shape exhibited in FAS DH, instead it possesses the liner shape found in PKS DH dimers. The MAS DH is composed as to be expected with a double hotdog fold with each hotdog fold contributing one of the amino acids needed to from the catalytic active site. The hydrophobic substrate tunnel is composed of both hotdog folds with the entrance located at the distal C terminal of the hotdog fold. [75]

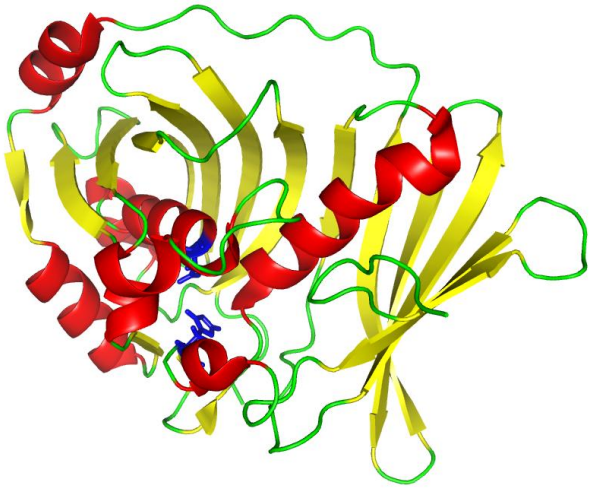


Figure 23 – Crystal structure of mycocerosic acid (MAS) DH from mycobacterium *smegmatis*. PDB: 5BP2.[75]

1.6 Project aims

Work carried out on mPKS DH domains has shown it is possible for DH domains to be excised from the whole PKS enzyme and that they can function independently from the rest of the enzyme.[6, 12, 13, 60, 62, 63, 65, 67, 86] The aim of the work described in this thesis is to investigate the programming of isolated SQTKS DH. The programming of the isolated SQTKS DH domain is to be probed using synthetic substrate mimics. No previous studies of iterative PKS DH domains have been reported and a key aim is to discover the chemo and stereo-selectivity of such domains.

The aims of the project are as follows:

- To synthesise a range of substrates which are distereomerically and enantiomerally pure. The synthesised substrates will be designed to probe the programming of the SQTKS DH domain.

- Optimise the production of the SQTKS DH domain protein to produced sufficient quantities of protein to carry out assays.

- Characterise the selectivity of the SQTKS DH domain and carry out steady state kinetic analysis of the substrate processed by the domain.

- Attempt to crystallise the DH domain to obtain a crystal structure. The active substrates will be docked into the active site to gain understanding of how the active site dictates the selectivity of the DH domain. In the absence of a crystal structures commercially available modelling programs will be used to create a model of the DH domain.

1.6.1 Squalestatin S1 (SQS1)

Squalestatins are secondary fungal metabolites produced by filamentous fungi and over the years many different squalestatins have been isolated and characterised.[87] All squalestatins comprise a highly functionalised 4,8-dioxabicyclo[3.2.1.]octane core with varying 1-acyl and 6-*O*-acyl side chains.

Squalestatin S1 (SQS1, also known as zaragozic acid**), 37**, was first isolated in the early 1990's from *Phoma* sp. C2932 (Glaxo) and an unidentified strain MF5453 (Merck), as a cholesterol lowering drug due to its potent inhibition of squalene synthase.[88, 89] When first discovered, the unusual structure of **37** raised questions about the biosynthesis of the carbon backbone and the origin of the high number of oxygen atoms.[90] Stable isotopic labelling studies showed **37** was composed of two polyketide chains.

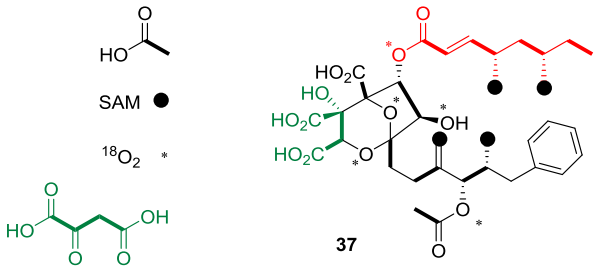


Figure 24 – Structure and origin of atoms in SQS1 **37**.

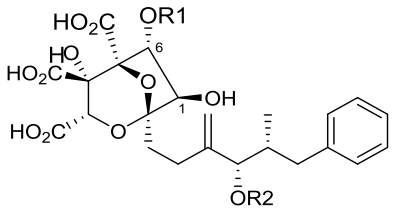
The main chain (shown in black) is a hexaketide, comprising a benzoate starter unit, thought to derive from phenylalanine, and 5 acetate-derived extender units.[90] The second chain (shown in red) is a tetraketide derived from an acetate starter and extender units. The distinctive bicyclic core structure is completed by the addition of 4 carbons deriving from oxaloacetate. Fungal PKS are known to possess a functional *C*Met domain which is able to transfer methyl groups from SAM. SQS1 contains 4 methyl groups, 2 on each chain, which have been installed by *C*Met domains.[90, 91]

1.6.2 Previous work

Previous work by the Cox group into the biosynthesis of SQS1 **37** resulted in the identification and isolation of the gene responsible for the production of the tetraketide side chain, *phpks1* which encodes squalestatin tetraketide synthase (SQTKS).[91]

Further work by Cox *et al*. identified the full biosynthetic gene cluster responsible for the production of SQS1 for the first time.[92] Full genome sequencing of the *Phoma* sp. C2932 (Glaxo) and an unidentified MF5453 (Merck) strain was carried out. Using the previously characterised *phpks1* gene it was possible to identify a putative SQS1 gene within the assembled genomes.

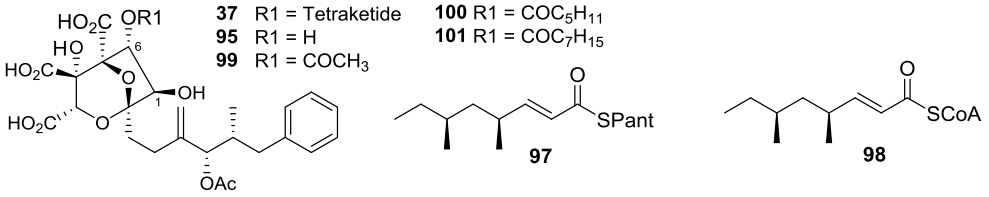
The link between the gene cluster and the biosynthesis of SQS1 was strengthened by growing MF5453 under producing and non-producing conditions and looking at the gene expression levels. Under producing conditions there was found to be a 10^2 to 10^5 fold increase in the expression levels of the genes in the putative SQS1 cluster.



37 R1 = tetraketide, R2 = $COCH_3$
94 R1 = H, R2 = H
95 R1 = H, R2 = $COCH_3$
96 R1 = tetraketide, R2 = H

Scheme 25 – SQS1 and related analogues.

The cluster was found to contain 2 PKS encoding genes. One is known to be responsible for the tetraketide side chain, therefore the other must encode the hexaketide chain. The hexaketide synthase contains an AT domain with suitable residues next to the active site to allow the selection of a benzoate starter unit. The cluster also contains a citrate synthase-like protein which is thought to be involved in the linking of the hexaketide to the oxaloacetate moiety.

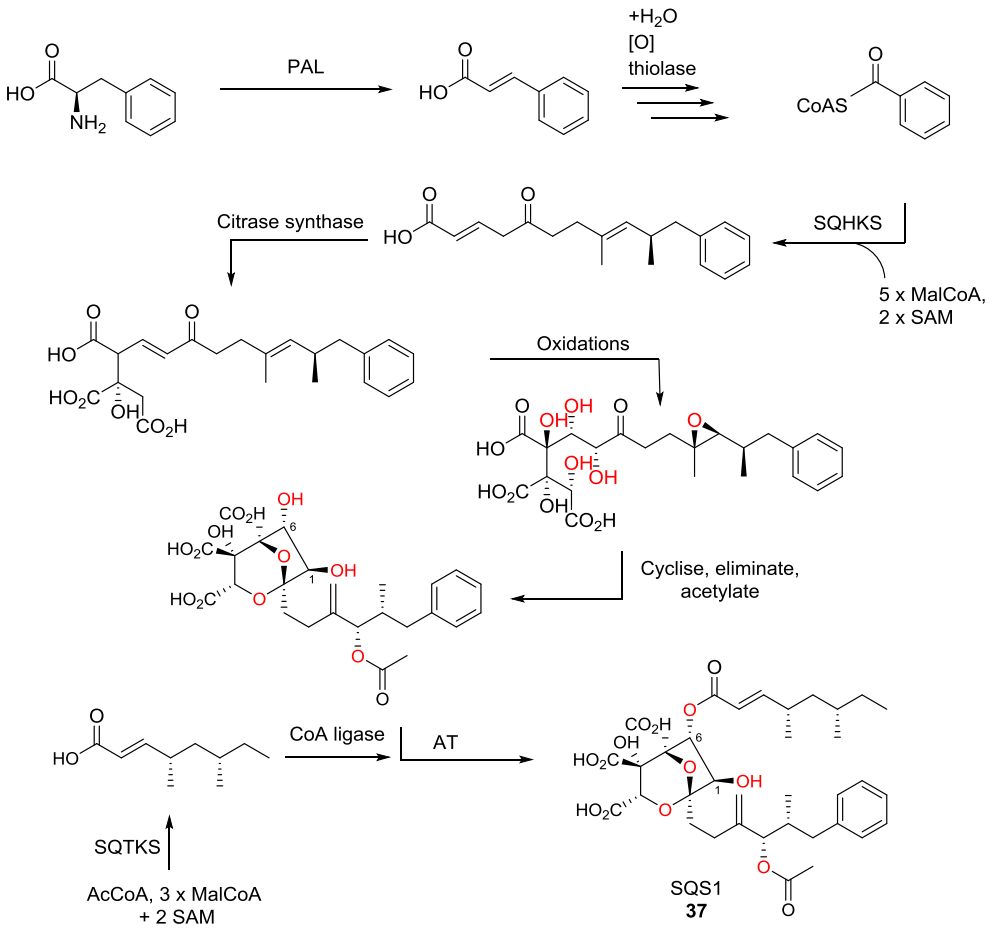


Scheme 26 – *In vitro* substrates and SQS1 product and analogues.

In order to gain an insight into the order in which acyl groups are added into the final compound, *phpks1* was knocked out. In the wild type, when grown under producing conditions SQS1, **37** is the main product, with trace amounts of **94**, **95** and **96**. In the *phpks1* knockout, there was no production of SQS1 **37**, but there is very high production of **95**. The results gained

from this experiment indicate that the transfer of the tetraketide is probably the final step of biosynthesis (Scheme 27).

The SQS1 gene cluster contains two putative ATs, and it was possible to isolate and express one of these AT domains. To investigate the selectivity of the AT domain, various possible substrates were used: **94**, **95** and **96**, were incubated in the presence of the AT domain and various acyl groups (**97** and **98**) and acetyl, hexanoyl and octanoyl CoAs. The assays showed that the AT domain was unable to transfer the tetraketide pantetheine substrates. When the AT was incubated with the tetraketide CoA substrate, SQS1 **37** was formed. In addition to transferring the tetraketide from its CoA substrate, the AT domain was also able to transfer acetate, hexanoate and octanoate from their corresponding CoAs to form **99**, **100** and **101**. The AT domain was unable to transfer acyl groups to the 12-alcohol in **94** or **96**. The results confirmed that this AT domain is responsible for transferring the tetraketide to form SQS1 **37**. The acetate must be transferred to *O*-12 at an earlier stage in the biosynthesis. From examining the gene cluster, undertaking gene knock-outs and isolating the one of the AT domains, the Cox group could propose the biosynthesis of SQS1 (Scheme 27)

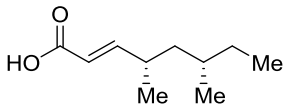


Scheme 27 – Proposed biosynthesis of SQS1 **37**.[92]

1.6.3 Squalestatin Tetraketide Synthases (SQTKS)

SQTKS is a type I, iterative, highly reducing fungal PKS, consisting of a single polypeptide of 2603 amino acids.[93] There are seven identifiable catalytic domains; KS, AT, DH, *C*-MeT, ER, KR and ACP. The KR domain is split into two parts; the catalytic KR and the structural ψKR. Currently, it has not been possible to obtain a crystal structure of the entire synthase or the isolated domains. Protein sequence alignment suggests that SQTKS has a similar structure to that found in FAS (Figure 3).

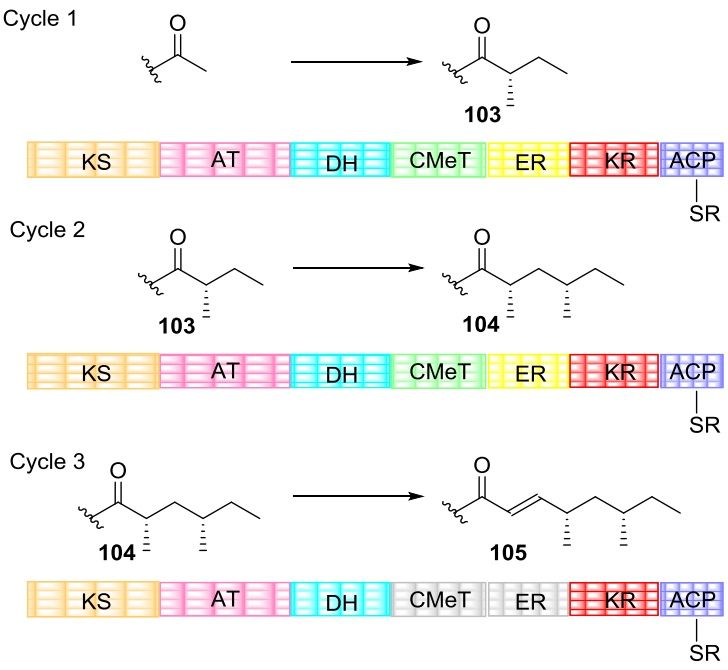
SQTKS is interesting for a number of reasons; it is currently the simplest known HR-iPKS with all modifying domains present and active. SQTKS also exhibits complex programming. Squalestatin tetraketide **102**, is biosynthesised from acetate starter and malonate extender units and methylation.[91, 93]



102

Figure 25 - Structure of squalestatin tetraketide **102**.

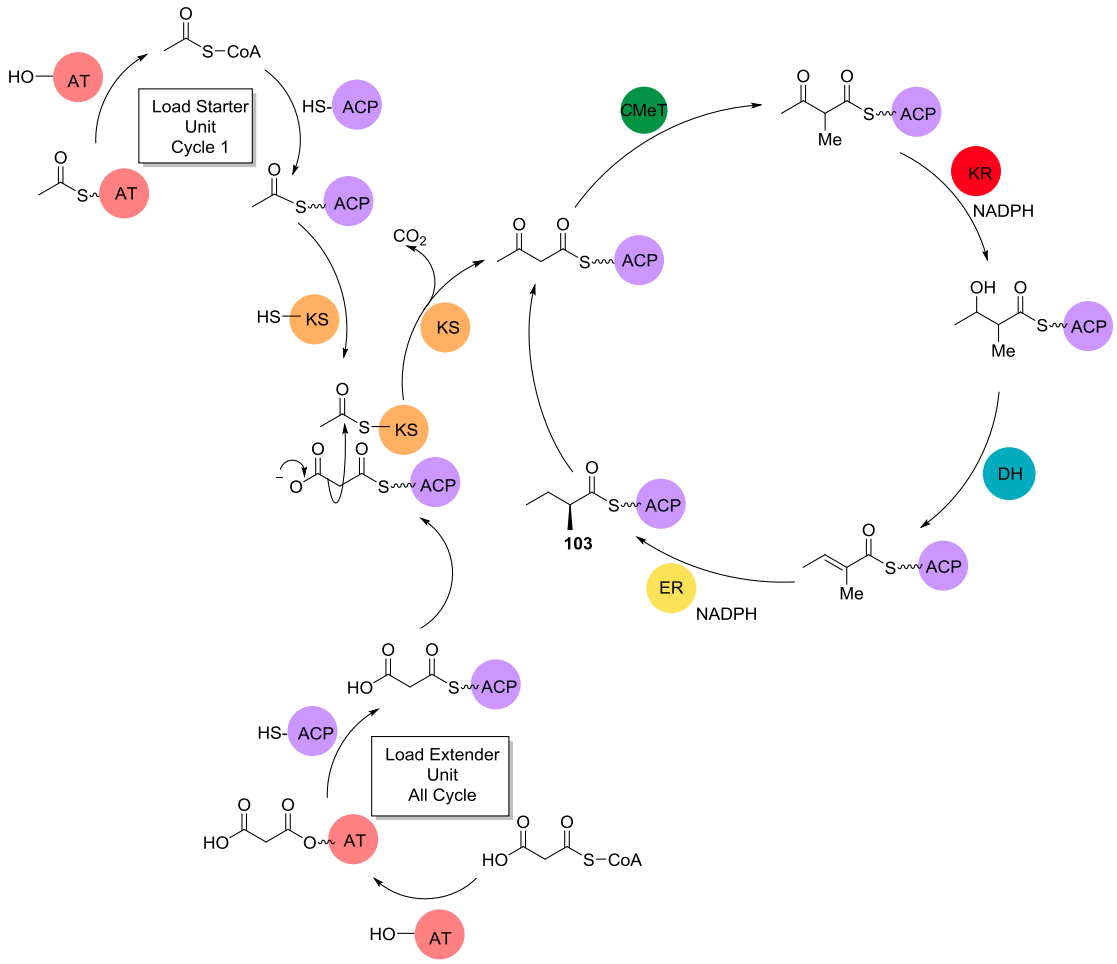
There are three rounds of chain extension and modification (Scheme 28). All domains are active in the first and second round of chain extension, resulting in a fully saturated diketide. In the third and final round of chain extension the *C*-MeT and ER domains are inactive. This results in the final product **102** lacking an α-methyl substituent and retaining the α,β-unsaturation. SQTKS must also contain some programming enabling control over the stereochemical outcome of certain steps in the biosynthetic cycle. As SQTKS lacks an obvious TE it is not clear how the polyketide is released.



Scheme 28 - Biosynthesis of squalestatin tetraketide.

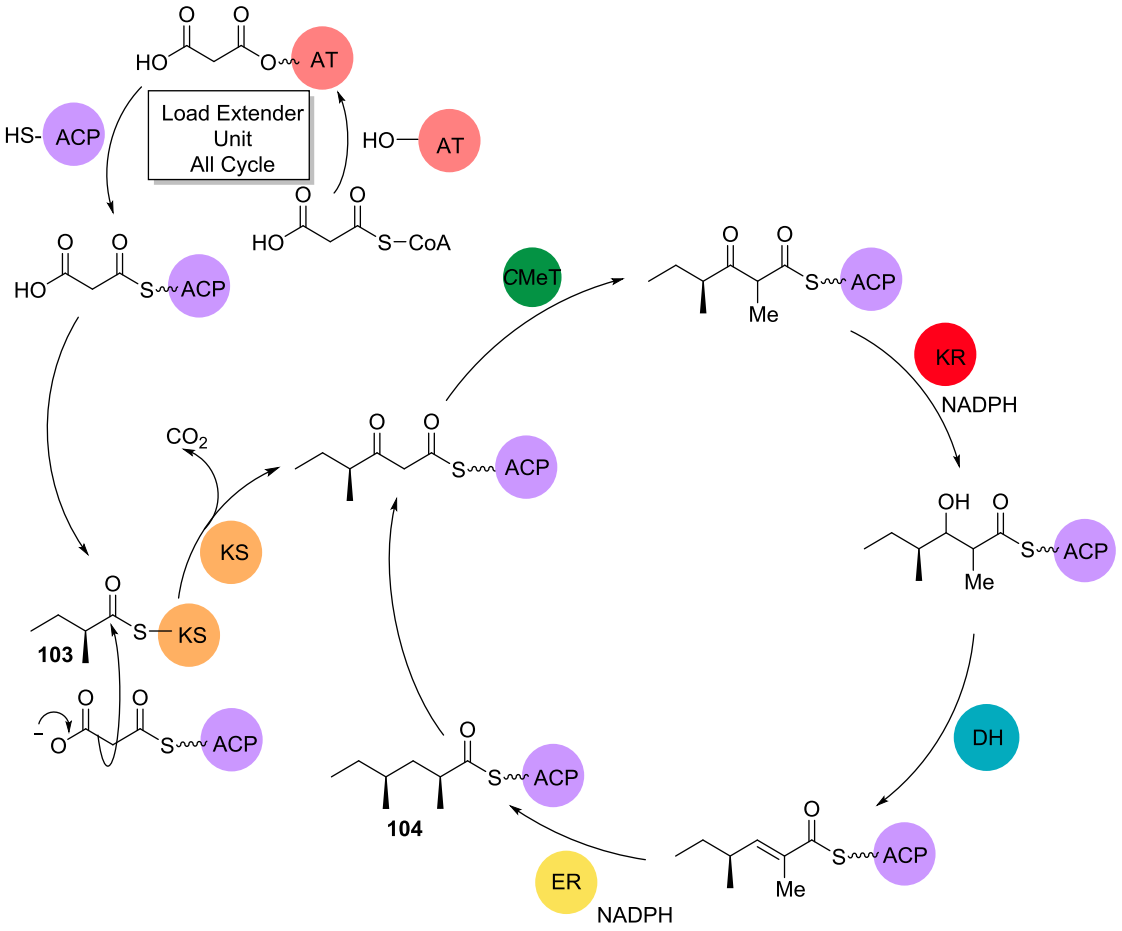
In the first biosynthetic cycle, the acetate starter unit is loaded onto the AT domain before being transferred onto the ACP and subsequently onto the KS domain (Scheme 29). The malonyl extender unit is then loaded on the ACP *via* the AT domain. Once both the starter and extender unit have been loaded, the KS catalyses the condensation of the two units *via* a Claisen condensation reaction to form an ACP bound β-keto thiolester.

The ACP bound thiolester is then presented to the *C*-MeT domain, where it undergoes α-methylation. The KR domain reduces the β-ketone group to a β-OH. The DH domain then eliminates water forming an α,β-unsaturated thiolester, which is then subsequently reduced by the ER domain resulting in a diketide thiolester **103**, which is fully saturated at the β-position.



Scheme 29- Cycle one of SQTKS biosynthesis.

The diketide, **103**, is then fed back into the catalytic cycle for round two of the chain extension (Scheme 30). A second malonyl extender unit is loaded and coupled to form a triketide. The triketide thiolester is then methylated by the *C*-Met, reduced by the KR, dehydrated by the DH and finally reduced by the ER to form the fully saturated product **104**.



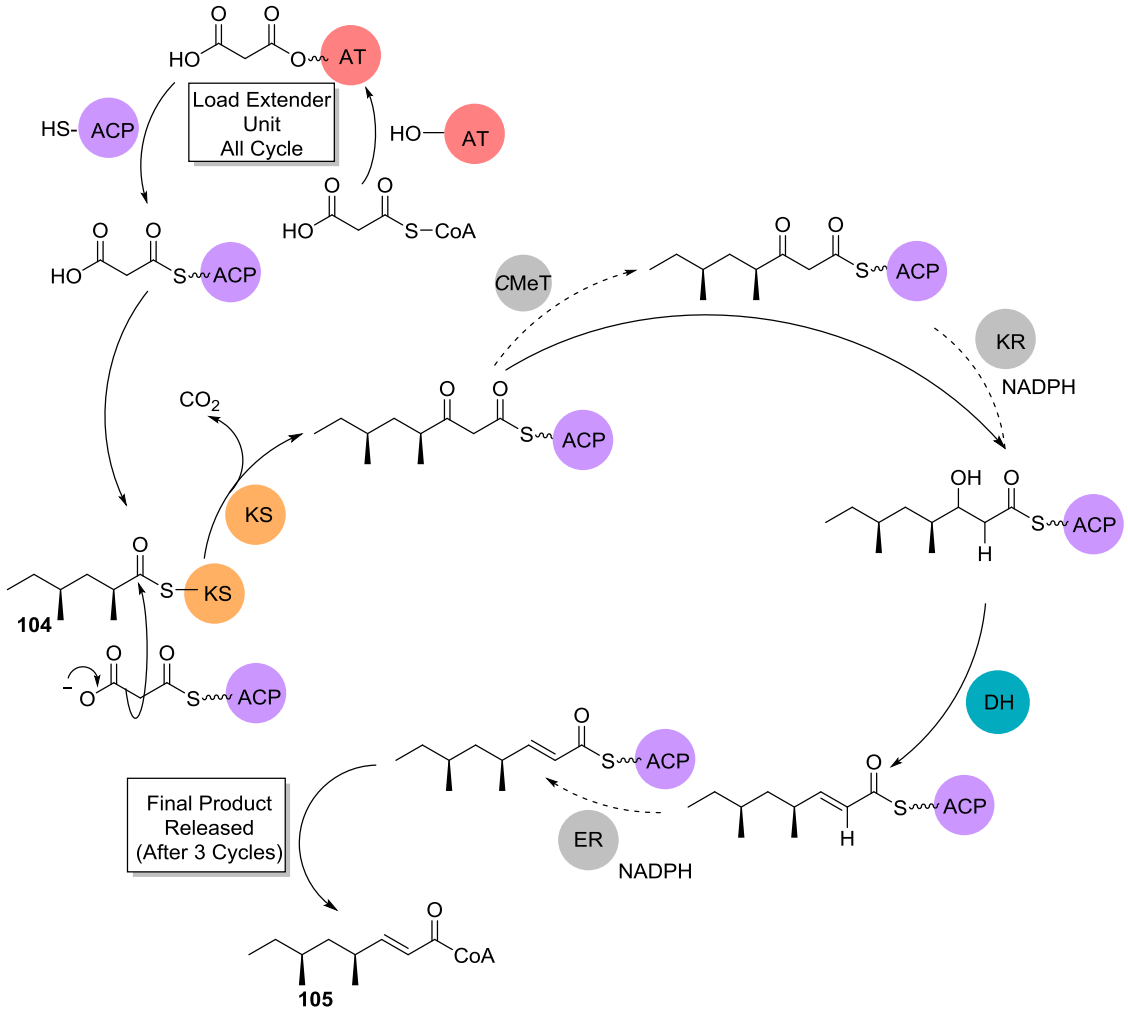
Scheme 30- Cycle two of SQTKS biosynthesis.

In the final round of chain extension **104** is fed back into the catalytic cycle and again condensed with a third malonyl extender unit to form a tetraketide (Scheme 31). In this cycle the *C*-MeT domain has been "switched off" and does not methylate in the β-position. The KR and DH domain are active resulting in an α, β-unsaturated thiolester. As with the *C*-Met domain, the ER domain is inactive and does not react. This results in the final product **105**, retaining the α,β-unsaturation. There is more than one stereochemical outcome possible for each stage of the biosynthesis, although only one final product is formed. All the possible outcomes for the diketide and triketide products are shown in Scheme 32. In the third cycle (Scheme 31) of chain extension the *C*-Met and the ER domain are inactive and the stereochemical possibilities have not been illustrated in this report.

The *C*-MeT, KR, DH and ER domain of SQTKS must contain a degree of programming in order to control the stereochemical outcome and ensure the correct product is produced. The ER domain dictates the stereochemistry of the final product in the first and second round of chain extension. The *C*-Met and DH and ER all affect the outcome of the final round of chain extension.

A number of questions arise regarding the programming of the SQTKS. Although the ER domain affects the final stereochemistry of the product, the stereochemical outcomes from each stage in the biosynthesis are important. The outcome from one domain is likely to affect the action of the subsequent domain. By discovering the selectivity of SQTKS DH, it will be possible to predict the product formed by the KR and the CMeT domain.

Currently it is known that the ER domain catalyses the reduction of the α,β-unsaturated thiolester resulting in methyl groups having *S* stereochemistry. It is not known what causes the ER domain to become inactive in the final round of chain extension. Currently work is under way in the Cox group to investigate the programming of the ER domain. The programming of the domain could be affected by a number of factors including chain length or methylation pattern (e.g. the lack of α-methylation). Current findings from ER studies show *Z* alkenes can also be a substrate for the ER domain,[94] this makes the understanding of the DH domain programming even more important.



Scheme 31- Cycle three of SQTKS biosynthesis.

1.6.3.1 Dehydratase domain of SQTKS

Previous work done within the Cox group[95] by David Ivison resulted in the purification of the SQTKS DH as a standalone protein and initial testing to characterise it structurally and chemically was carried out.

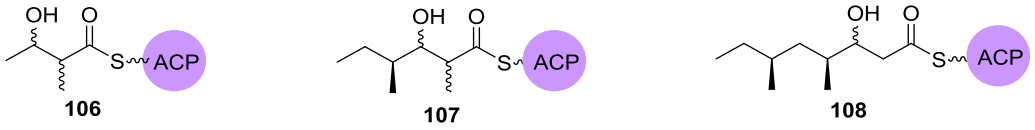


Figure 26 - DH Natural substrates.

There are three natural substrates for the DH domain. The diketide **106**, triketide **107** and the tetraketide **108**. The diketide **106** has four possible stereoisomers **109a - d** (Scheme 32). The triketide **107** has eight possible stereoisomers (4 are shown in Scheme 32 **110a - d**), although the 4-methyl must be as shown in **107**.

Due to the number of possible substrates for the triketide **107**, previous work was carried out using the diketide substrate. A diketide DH substrate mimic, 3-hydroxy-2-methylbutanoyl-SNAC **111**, was synthesised as a mixture of diastereomers and assayed. There are two possible products from the reduction of **111**, tigloyl-SNAC **112**, (*E*-enone) and angelic-SNAC **113**, (*Z*-enone).

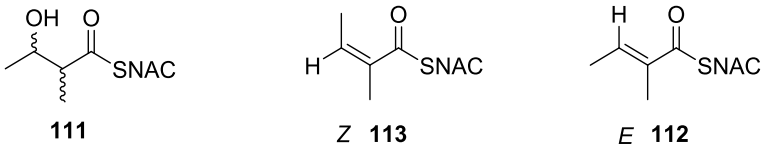
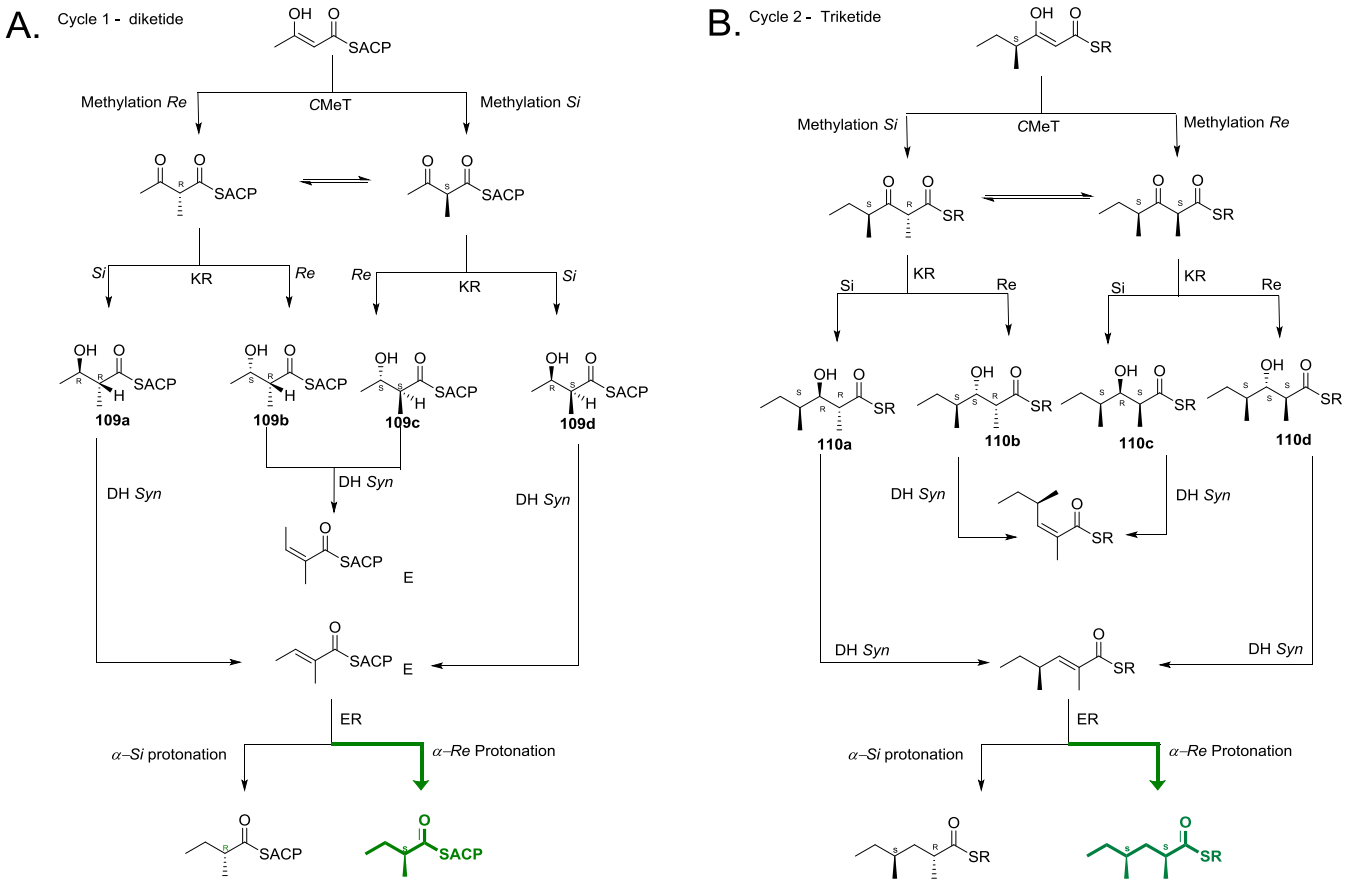


Figure 27 - 3-Hydroxy-2-methylbutanoyl-SNAC **111**, tigloyl-SNAC **112,** and angelic-SNAC **113**.

An assay was carried out by incubating the DH protein at 37 °C in the presence of racemic 3-hydroxy-2-methylbutanoyl-SNAC **111** overnight, alongside a control assay. When analysed by LC-MS, a new peak was observed in the enzyme assay but not in the control assay. The peak was found to have a *m/z* 224, which corresponds to the formation of α,β-unsaturated thiolester-SNAC [M] Na^+. In order to fully identify the new peak at 5.1 mins, the assay was repeated and the assay mixture extracted with deuterated chloroform and analysed by NMR which confirmed the formation of tigloyl-SNAC **112**.

The SQTKS DH catalysed reaction is predicted to proceed *via* a *syn*-elimination mechanism. *Syn*-elimination reactions occur *via* an E1cB mechanism with the formation and collapse of an enol(ate), resulting in the loss of water (Scheme 13).



Scheme 32 - A. Flow diagram showing all stereochemical possibilities of diketide β-processing. The product produced by SQTKS after the first cycle is highlighted in green. **B.** Flow diagram of SQTKS domain reaction and all possible stereochemical outcomes for triketide product. The product with the correct stereochemistry at the end of the second cycle is highlighted in green.

Currently the stereochemistry of the actual DH substrate is unknown. There are four possible stereochemical isomers for the diketide substrate; (*2R, 3R*)-3-hydroxy-2-methylbutanoyl-SNAC **114**, (*2S, 3S*)-3-hydroxy-2-methylbutanoyl-SNAC **115**, (*2R, 3S*) - 3-hydroxy-2-methylbutanoyl-SNAC **116** and (2*S*, 3*R*) - 3-hydroxy-2-methylbutanoyl-SNAC **117**.

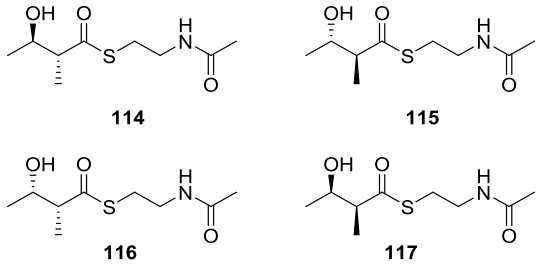
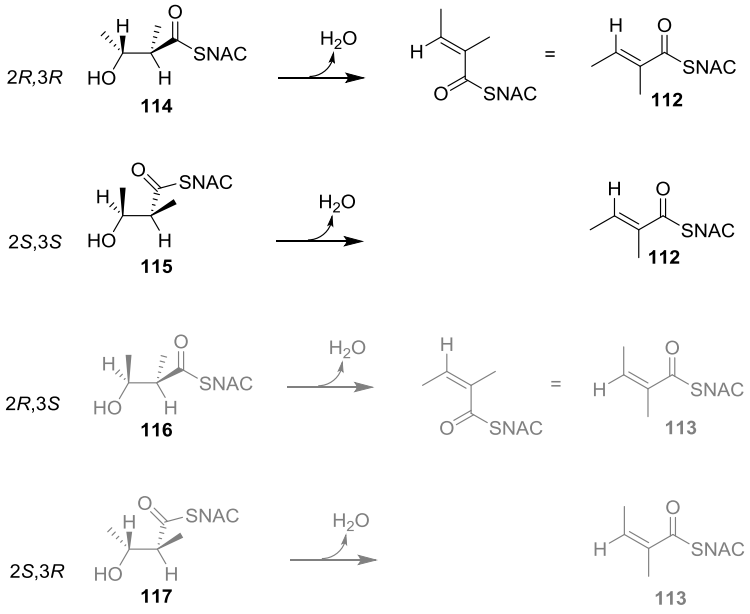


Figure 28 - Natural substrate mimics.

Assuming that the α,β unsaturated thiolester forms *via syn* elimination and that tigloyl-SNAC **112** is the only product, it can be predicted that (*2R,3R*)-**114**, or (*2S,3S*)-**115**, can be used as a substrate for the SQTKS DH. *Syn*-elimination of the *(2R,3S*)-**116** and *(2S,3R*)-**117** substrates would produce the angelic-SNAC **113** (Scheme 33). In order to test this hypothesis a small library of substrate mimics will be synthesised which are diastereomeric and enantiomerically enriched in order to probe the selectivity of the SQTKS DH domain.



Scheme 33 - Stereochemical outcomes of *syn*-elimination of the four possible diastereomers of 3-hydroxy-2-methybutanoyl-SNAC.

Chapter 2 - Synthesis of DH substrate mimics

2.1 Introduction

The stereochemistry of the final product of SQTKS, squalestatin tetraketide **108**, is a result of the programming present in the iPKS. Each domain present in the iPKS will possess a degree of programming. Work by Douglas Roberts and Christoph Bartel within the Cox group has probed the programming of the SQTKS ER domain.[94, 96] Their work has shown that the isolated ER domain is able to tolerate and turn over a wide range of substrates regardless of the methylation pattern, chain length or double bond geometry (Figure 29). In order to probe the programming of the SQTKS ER domain, a library of ER substrates was synthesised. The different stereochemistry of the substrates was designed to interrogate the programming of the active site and shed light on the selectivity of the SQTKS ER domain.

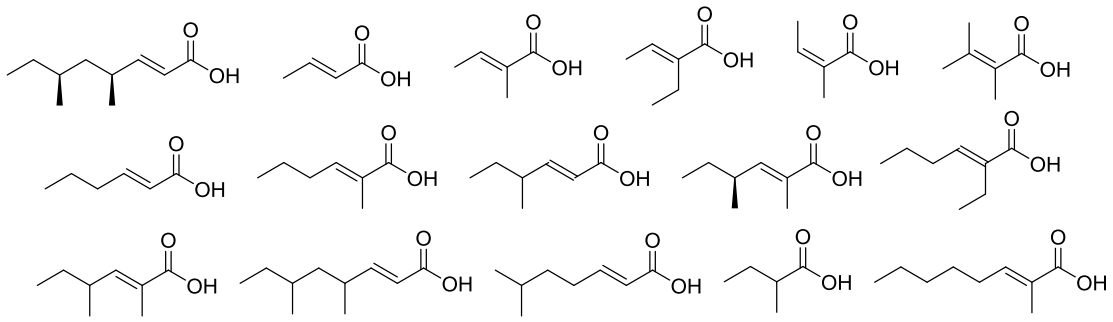


Figure 29 -Substrates synthesised by Roberts *et al.* designed to be attached to SNAC and pantetheine thiolesters.[94]

Several PKS domains have been interrogated with substrate mimics in order to gain an understanding of their programming.12, 13, 17, 60, 64 During PKS biosynthesis, the substrate is bound to an ACP and ideally, assays to investigate the programming within SQTKS DH would be carried out using synthetic analogues of natural substrates attached to the ACP. Within the Cox group, work was carried out by David Ivison to produce SQTKS-ACP as a soluble protein.95 SQTKS-ACP was cloned into a pOPINF vector, which was subsequently transformed into E. coli for expression. There was no expression of ACP observed under any of the experimental conditions trialled. Also although it is possible to produce acyl-ACPs,13, 71, 97, 98 they are not ideally suited to kinetic based assays, due to their large mass, resulting in large quantities being required. Also the ACP must be cleaved from the acyl product before analysis can be done.

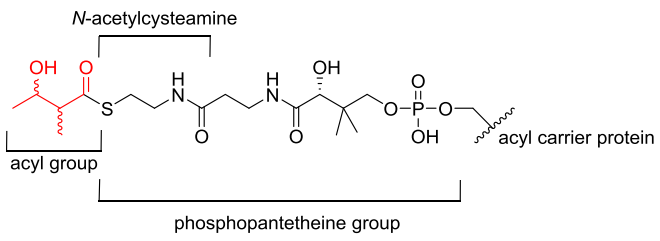
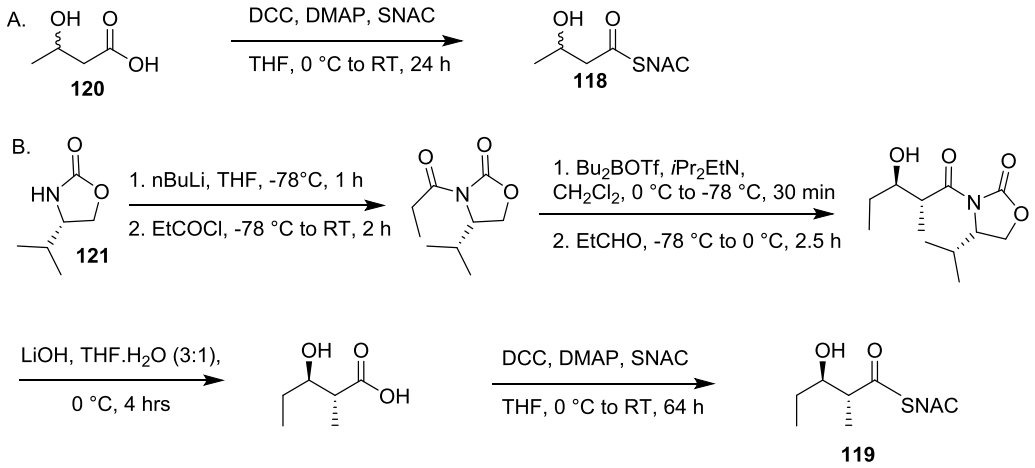


Figure 30 - Polyketide chain, shown in red, attached to a phosphopantetheine ACP.

Acyl substrates are attached to the ACP *via* a phosphopantetheinyl group (Figure 30) which acts as a long flexible arm, allowing the movement of the polyketide chain from domain to domain within the PKS module.[21] Substrate mimics have been synthesised which, rather than utilising the native ACP, are attached to only the phosphopantetheinyl group or shortened even further to an *N*-acetylcysteamine (SNAC) thiolester.

A number of PKS DH domains have had their substrate specificity investigated.[6, 13, 60, 62, 65-67] The majority of DH domains investigated are from type I mPKS, with the exception of curacin **93** which is produced by a PKS-NRPS.[17] The substrate specificity of the DH domains was probed using SNAC, phosphopantetheinyl and ACP based substrates. Typically substrate mimics have been produced using synthetic organic chemistry, utilising chiral auxiliaries to control the stereochemistry of the products formed. Substrates have also been produced enzymatically.[13]

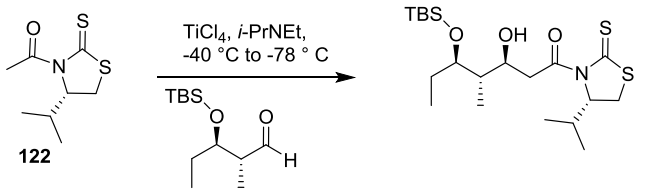
Work by Vergnolle *et al*. utilised SNAC substrates when investigating the formation of *cis* double bond formation in borrelidin **50**.[60] In order to probe the selectivity of these domains, a range of diketides substrate mimics **118** and **119** were synthesised. The non-methylated diketide mimics for BorDH2 were simply synthesised by coupling the SNAC to the required 3-(*S*) or (3R)-hydroxybutanoic acid **120** in the presence of DCC and DMAP (Scheme 34, A). For the methylated diketide substrate mimics Vergnolle *et al*. utilised oxazolidinone auxiliary chemistry to synthesise the four stereoisomers of 3-hydroxy-2-methylpentanoyl SNAC (Scheme 34, B).[60]



Scheme 34 – Synthesis routes used by Vergnolle *et al.*[60] to produce borrelidin substrate mimics; **A**, diketide mimic synthesis; **B**, methylated diketide mimic synthesis.

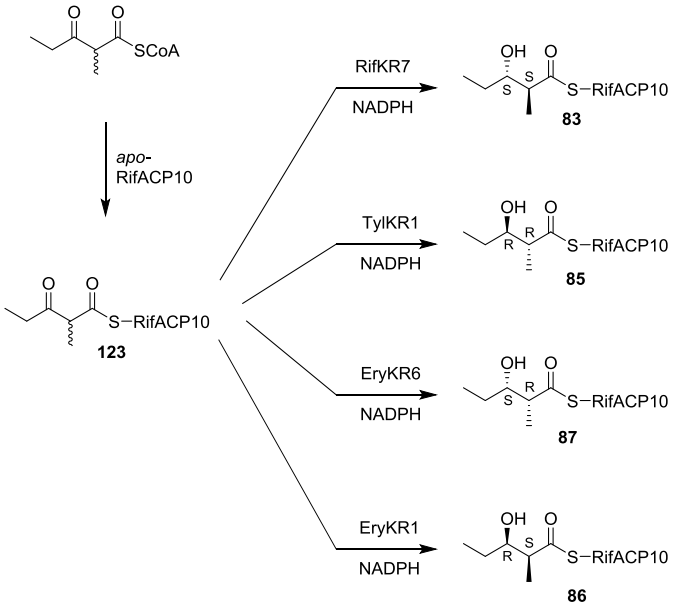
The DH domain from pikromycin, PikDH2, is another type I mPKS DH which has had its selectivity probed.[65, 67] The substrate mimics synthesised by Li *et al.* were also based on simple diffusible SNAC thiolesters.[65, 67] Although there was one difference in the design of the triketides; one or two methylene spacers had been inserted in between the carbonyl and the

SNAC moiety to remove the risk of internal lactonisation occurring in the substrate mimics. As was the case with the work by Vergnolle *et al*., Li *et al.* which utilised chiral auxiliaries **121** to control the stereochemical outcome aldol reactions when synthesising substrate mimics (Scheme 35). Although rather than using oxazolidinone auxiliaries, Li *et al.* used a thiazolidinedione auxiliary **122**, also known as Masamune auxiliary.



Scheme 35 – Aldol reaction carried out by Li *et al.* using thiazolidinedione auxiliary.[67]

Wu *et al*. also investigated the formation of olefin in pikromycin, but rather than investigating the DH domain PikDH2 in isolation, the entire Pik2 module was used.[65] Wu *et al*. synthesised DH diketide substrates based on simple diffusible SNAC thiolesters. The substrate was synthesised using Evans oxazolidinones to control the stereochemistry of the acyl substrates synthesised.[65] In order to study the terminal rifamycin DH, Rif10DH, Gay *et al.* produced all four stereoisomers of 2-methyl-3-hydroxypentanoyl-RifACP10.[13] Previous work by the Keating-Clay group had shown it was possible to employ mPKS KR domains as biocatalyst to make chiral diketide mimics **83** and **85-86**.[99] The stereoisomers were formed by incubating (2*RS*)-2-methyl-3-oxopentanoyl-RifACP10 **123** with KR domains from rifamycin, (RifKR7), tylosin (TylKR1), erythromycin , (EryKR1 and EryKR6), (Scheme 36).



Scheme 36 – Chemoenzymatic synthesis of diketide substrate mimics using various PKS KR domains.[13]

Although work by Piaskecki *et al*. showed that it was possible to scale up the chemoenzymatic reactions allowing over 100 mg of product to be isolated,[99] this approach for substrate mimic

synthesis was not suitable for our studies. In order to use this approach, it would require expression of SQTKS ACP as a soluble isolated protein and currently this is not possible,[95] It would also require access to the KR domains used in the study which are also not easily available.[99] The cost of obtaining methylmalonyl-CoA and NADPH is another deterrent to using chemo enzymatic reactions to synthesis substrates on a useable scale.

Another factor to take into account with the chemoenzymatic route, is substrate fidelity. Changes in a domain's selectivity can occur when the ACP is changed from a native to a non-native ACP.[13] This was observed during the investigation into the selectivity of rifamycin DH (Rif10DH) and is discussed in Chapter 1, Section 1.5.7. It is unknown if the stereochemical outcome of the enzymatic reactions will be retained if the ACP from SQTKS is used.

When designing substrate mimics for this study, it was important to consider how the assays were to be carried out. A range of assay conditions and sample work-ups have been reported by others. The simplest assay conditions were used by Vergnolle *et al.*[60] and Li *et al.*[67] Vergnolle incubated the borrelidin substrate mimics (12 mM) with 25 mM HEPES, 100 mM NaCl and up to 8 mg mL^{-1} of either the BorDH2 or BorDH3 protein.[60] The reactions were incubated at 37 °C for 16 hours. The work-up of the assay involved carrying out an extraction (2 × 1 mL) using EtOAc. The samples were then evaporated to dryness and resuspended in 50 µL of isopropanol and subsequently analysed by LC-MS.

Li *et al*. also used a simple assay.[67] The substrate was incubated with the buffer and PikDH2 protein for 40 minutes at 25 °C. At certain time points, 5 µL samples were taken and quenched in 495 µL of 1:1 MeCN and buffer. This resulted in rapid protein precipitation. The samples were then centrifuged to precipitate the protein and 60 µL was added to a HPLC vial for analysis by LC-MS/MS. Not only were Li *et al*. able to carry out assay of PikDH2 using SNAC based substrates, they were also able to carry out kinetic assays, allowing the kinetic characteristics of the PikDH2 domain to be probed. The reaction conditions were the same for the kinetic assays, although the assays were quenched after only 15 minutes.

The aim of the assay carried out by Wu *et al.* was to incubate the diketide substrate, with Pik module 2 in the presence of malonyl-CoA and NADPH and observe the formation of the unsaturated triketide product.[67] This resulted in a more complicated assay and sample processing. The Pik module 2 had been modified so a functional TE domain was appended downstream of the ACP, in order to release the product as a carboxylic acid rather than as a SNAC derivative. The formation of product was initially detected using TLC phosphor imaging. The structure and stereochemistry of the product was confirmed by ^{1}H NMR and GC-MS. In order to achieve this a preparative scale assay was required, followed by extraction and purification of the product. In order for the products to be analysed by GC-MS, derivatisation had to be carried out using bis(trimethylsilyl)trifluoroacetamide (BSFTA).

These findings suggest that using synthetic organic chemistry to prepare the required substrate is the most reliable route. There is a strong literature precedent for the use of oxazolidinone auxiliaries to control the outcome of aldol reactions, Evans chemistry can be used for accessing the *anti* stereoisomer, but another route to accessing the *anti* stereoisomers is Fráter-Seeback alkylation.[100]

The method by which the samples are to be analysed dictated how the samples will need to be prepared. In order for SNAC mimics to be analysed by GC-MS, they require extraction from the assay mixture, purification and derivatisation, which can be time consuming and possibly lead to the introduction of errors, if kinetic data was to be collected in this fashion.[13, 65]

The work carried out by Li *et al*. gave the best grounding when designing the substrate mimics and assay protocol for the investigation into SQTKS.[67] The study utilizes simple substrate mimics, which are easy to synthesise with a high level of purity. In order to test the selectivity of SQTKS it is vital to be able to synthesise each stereoisomer in a pure form. The assays used are simple and quick to set up, with minimal work needed to quench and work up the assay samples. The design of the substrate and assay conditions allows the samples to be analysed by LC-MS, which obviates the need for derivatisation, which is required for analysis *via* GC-MS. Also Li *et al.* were able to carry out kinetic analysis of the PikDH2 domain. Currently it is not known how selective the SQTKS DH domain is towards certain substrate mimics. Kinetic analysis of the substrates would allow further insight into the efficiency of the DH enzyme.

2.2 Natural substrates

As discussed in Chapter 1, Section 1.6.3.1 there are three natural substrates for the SQTKS DH domain; the diketide **106**, the triketide **107**, and the tetraketide **108** (Figure 31). The diketide **106** has four possible stereo-isomers **114**, **115**, **116** and **117** and the triketide **107** has eight possible stereoisomers.

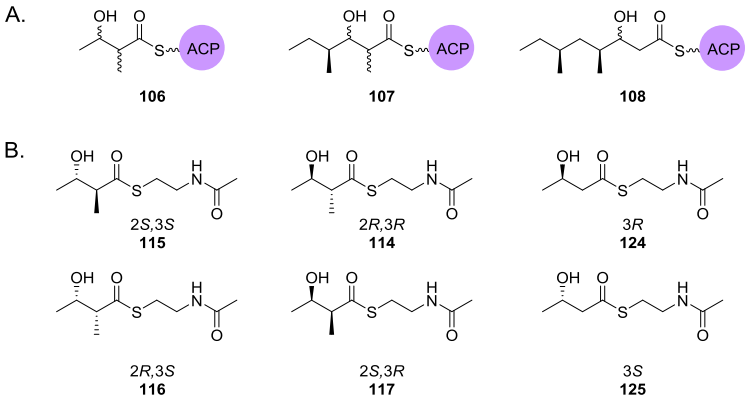


Figure 31 - A, Natural SQTKS DH substrates **106**, **107** and **108**; **B**, natural (**114**, **115, 116** and **117**) and non-natural (**124** and **125**) SQTKS DH diketide substrate mimics.

Due to the number of possible triketide substrates, it was decided to initially investigate the selectivity of the SQTKS DH domain towards diketide substrates **114-117**. There are four

possible stereochemical isomers for the diketide substrate and to make the synthesis as simple as possible it was decided that the acyl substrate mimics would be made as *N*-acetylcysteamine (SNAC) thiolesters. Non-natural substrates **124** and **125** were also synthesised to assess the importance of the methyl group in substrate recognition.

2.3 Synthesis plan for diketide substrate mimics

2.3.1 Synthesis of *syn* diastereomers

The synthesis of enantioenriched compounds using oxazolidinone methodology is well documented.[101-103] This methodology has been employed to make enantioenriched DH diketide substrate mimics in a number of DH studies.[60, 65, 67, 99, 102, 104] Oxazolidinone chemistry, also known as Evans auxiliary chemistry, offers a short route (four steps) to reach the desired compounds. A range of Evans auxiliaries **121**, **126-128** are commercially available and various reaction conditions which allow the synthesis of all four stereoisomers have been described in the literature.

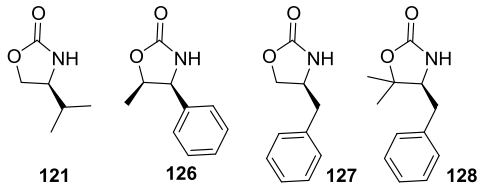
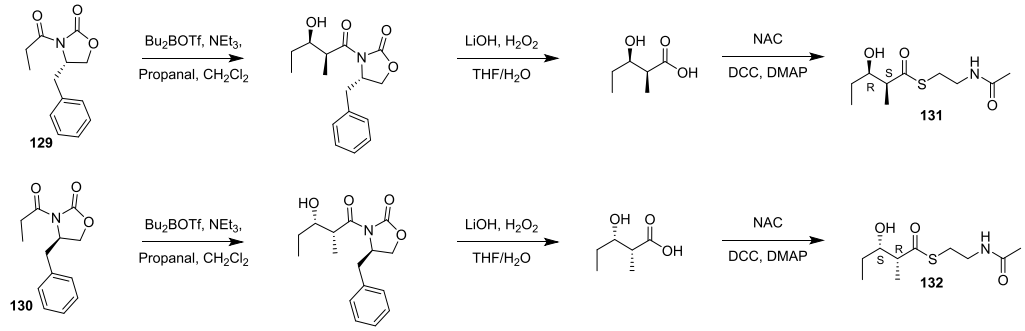


Figure 32 – Examples of commercially available oxazolidinone auxiliaries.

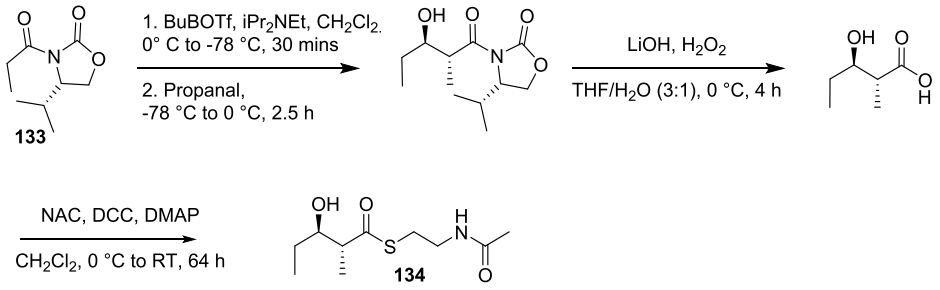
The Evans auxiliary, (*S*)-4-benzyl-2-oxazolidinone, **127**, has been used to synthesise SNAC related substrates.[102, 105] Staunton *et al.* used the acylated auxiliary **129**, as well as its enantiomer, **130**, during the synthesis of a range of diketide analogues of DEBS intermediates (Scheme 37).[102] (*S*)-4-benzyl-2-oxazolidinone **129** was used to prepare the (2*S*,3*R*)-substrate **131**, whereas (*R*)-4-benzyl-2-oxazolidinone **130** was used to access the (2*R*,3*S*)-substrate **132.** Keating-Clay employed an identical route when synthesising a range of diketide standards for his work investigating the programming of KR domains.[99]



Scheme 37 – Approach of Staunton *et al.* to prepare diketide substrate mimics using (*S*)-4-benzyl-2-oxazolidinone and (*R*)-4-benzyl-2-oxazolidinone.[102]

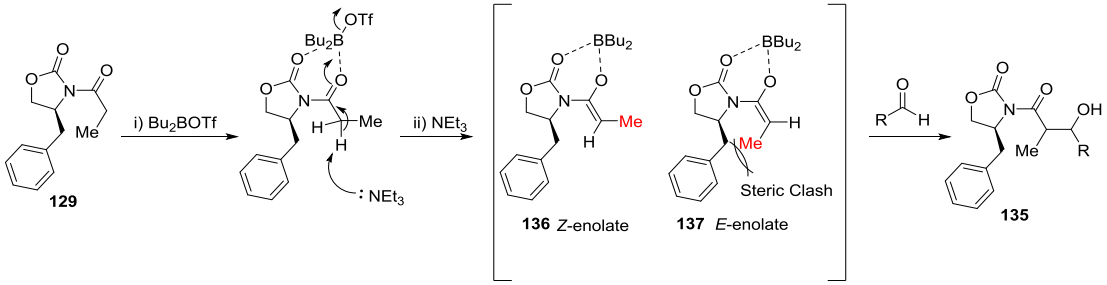
It is possible to use Evans auxiliary chemistry to access the *anti* diastereomers as well as *syn* diastereomers by altering the base used during the aldol reaction. The auxiliary (4*R*)-3-

propionyl-4-isopropyl-2-oxazolidinone **133** was used by Harris *et al.*[102] It was also used by Wu *et al*. when synthesising standards for the study of pikromycin DH **134** (Scheme 38).[65]



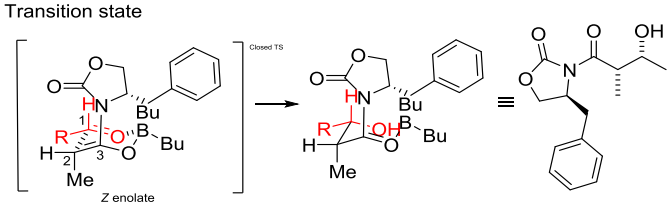
Scheme 38 - Synthetic route used by Wu *el al.* using the (4*R*)-3-propionyl-4-isopropyl-2-oxazolidinone.[65]

Chiral auxiliaries, such as oxazolidinones, control the stereochemical outcome of reactions in a number of ways depending on the type of reaction.[106] In aldol reactions, (Scheme 39), where **129** is transformed to **135**, oxazolidinones achieve stereochemical control by influencing the transition state of the reaction (**136** and **137)**.



Scheme 39- Formation of chiral enolate.

The use of oxazolidinones enables a chiral enolate to be formed by complexing to the butyl boron triflate (Scheme 39) and inducing the formation of the *Z*-enolate **136**. The *Z*-enolate is favoured, because the *E*-enolate **137** would result in greater alkylic strain between the methyl group (shown in red) and the benzylic group of the Evans auxiliary.

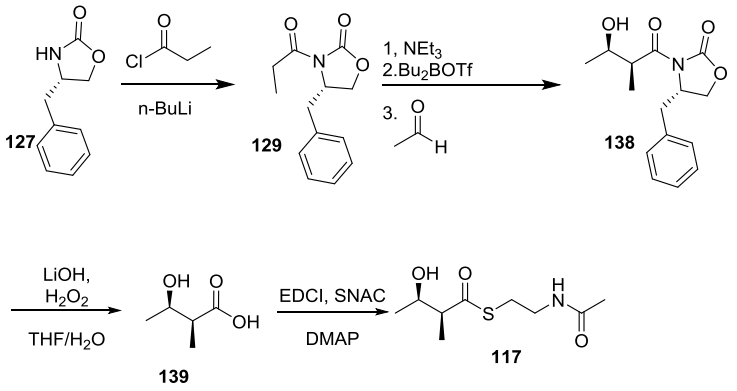


Scheme 40 - Stereochemical outcome of Evans aldol reaction under boron triflate conditions.[105]

Once the enolate has been formed and an aldehyde (red in Scheme 40) added, the boron complex forms a closed transition state in a pseudo chair conformation (Scheme 40).[107] In order for the chair conformation to form, the Evans auxiliary must orientate with the 5 membered ring and the methyl groups in pseudo axial positions. The orientation of the aldehyde can allow the

hydrogen to take either an axial or equatorial position, but the 1,3-diaxial interaction between the aldehyde and the oxazolidine forces the hydrogen onto the axial position and the R group to take the equatorial position.

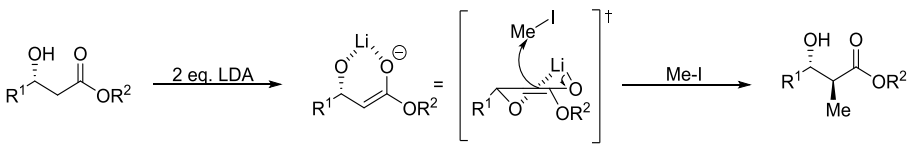
The synthetic route (Scheme 41) utilizes oxazolidinone chemistry to install the *syn* stereocenters in substrate **116** and **117**. It was decided to use (4*S*) and (4*R*)-4-benzoyl-2-oxazolidinones as there was a literature precedent for use of this Evans auxiliary.[99, 102] Evans auxiliaries have been reported to result in high stereoselectivity. Danda *et al*. reported diastereomeric ratio (d.r.) of 98:2 in favour of *syn* adducts.[108] Hayashi *et al*. also reported high selectivity for *syn* adducts, d.r. 100:0, when the aldol reaction was carried out using NEt_3 and Bu_2BOTf conditions to give **138**.[107] Once the aldol reaction had taken place, the Evans auxiliary can be cleaved by LiOH and H_2O_2 to give the acid **139**. The formation of **117** is achieved by coupling the acid **130** to SNAC in the presence of EDCI and DMAP.



Scheme 41 - Plan of (2*S*, 3*R*)-3-hydroxy-2-methylbutanoyl-SNAC **117** synthesis.

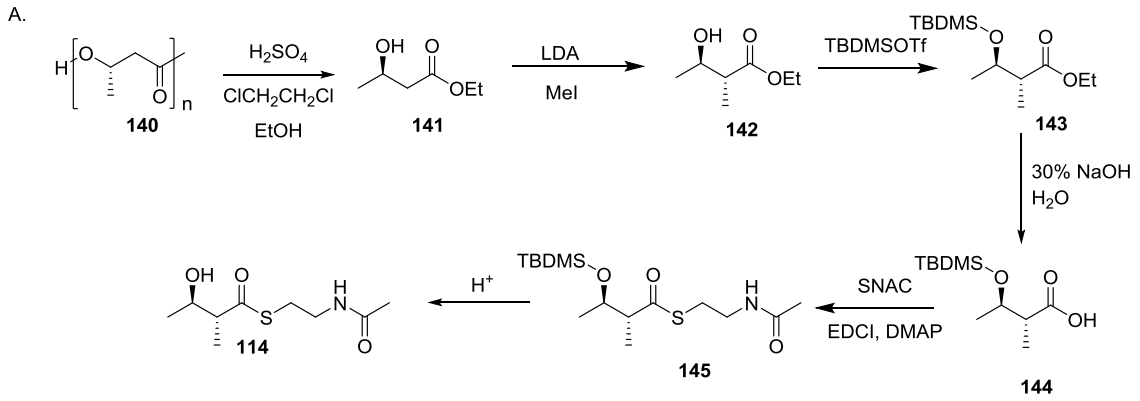
2.3.2 Synthesis of *anti* diastereomers

It is possible to produce the *anti* diastereomers *via* alternative methodology. Fráter-Seebach alkylation allows the diastereoselective alkylation of β-hydroxyl esters.[100] A strong base such as LDA is added to the β-hydroxyl ester, this deprotonates the hydroxy group and an α-proton allowing enolate formation (Scheme 42). The lithium ion co-ordinates to both of the oxygens, this causes one face of the enolate to be shielded, therefore the alkylation occurs on the least hindered face. In Fráter-Seebach alkylation, the least hindered face is the opposite face to the chiral β-hydroxyl group.[100] Fráter *et al*. reported the diastereoselectivity of the reaction to be 95:5 (19:1).



Scheme 42 - Fráter-Seebach alkylation.[100]

The synthesis route shown in Scheme 43 was proposed using Fráter-Seebach alkylation to introduce the desired *anti* stereochemistry in substrates **114** and **115**. Polyhydroxybutyrate **140** is a cheap, readily available starting material, which is readily converted to ethyl *R*-hydroxybutanoate **141**.[109]



Scheme 43 - Plan of (2*R*, 3*R*)-3-hydroxy-2-methylbutanoyl-SNAC **114** synthesis.

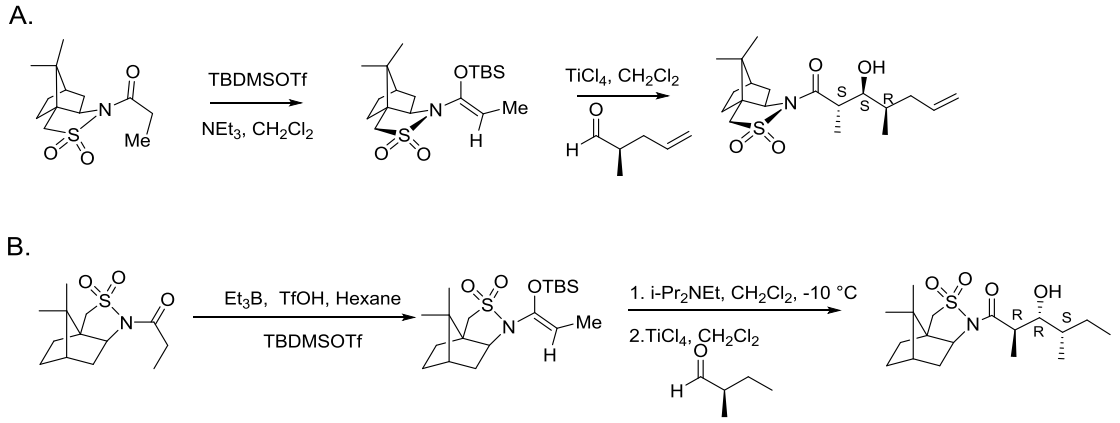
After the installation of the α-methyl group, the synthesis can proceed in a similar fashion to the planned route to the (*2S*,3*R*)-diastereomer.[102, 110, 111] It is also possible to synthesise the enantiomer (2*S*, 3*S*)-3-hydroxy-2-methylbutanoyl-SNAC **115** substrate using the same chemistry as described in Scheme 43. Instead of using *R*-hydroxybutanoate **141** the opposite enantiomer, *S*- hydroxybutanoate, can be used.

2.4 Synthesis plan for triketides substrate mimics

In order to further probe the stereoselectivity of the SQTKS DH domains, triketide substrate mimics could also be synthesised. The SQTKS DH triketide substrate has 8 possible stereoisomers, (Chapter 1, Section 1.6.3.1.). However as the stereochemsitry of the 4-position is known to be *S* from the structure of the tetraketide **108**, this allows the number of possible triketide substrates to be reduced to 4, **110a-d.**(Scheme 32).

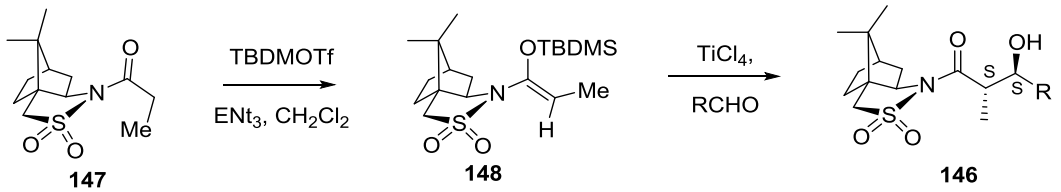
Scheme 44 – Structure of the final product, squalestatin tetraketide **108** and the triketide substrate **107**. The carbon atoms in the final product which arising from the triketide highlighted in red.

A literature search showed no evidence of Evans auxiliaries being used to synthesise similar compounds to **107** but camphorsultam directed aldol reaction had been used.[112, 113] For example Sang *et al*. utilised this reaction in their total synthesis of rakicidin A (Scheme 45, A).[112] McKillican *et al*. also carried out an asymmetric aldol reaction during their synthesis of avermectins (Scheme 45, B).[113]



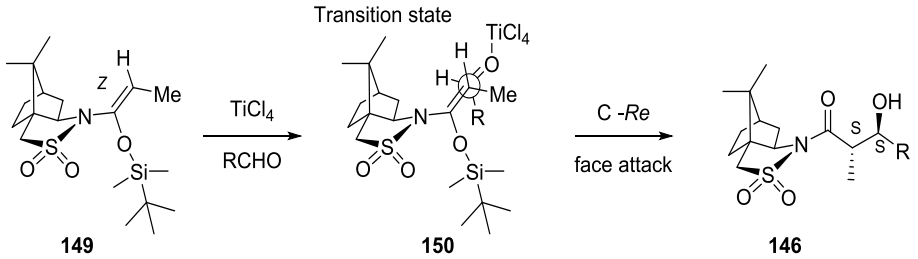
Scheme 45 - Synthetic routes utilizing camphorsultam directed aldol reactions to synthesise **A,** rakicidin A; **B**, avermectins.

The camphorsultam directed aldol reaction was developed by Oppolzer *et al*. in the early 1990s.[114, 115] It is possible to synthesis *syn* and *anti* aldols using boranesultam directed conversions by varying the conditions of the reaction. In order to achieve the 2,3-*anti* stereochemistry in **146** the propionylsultam **147** is first treated with TBDMSOTf and NEt_3, this results in the formation of a *Z*-silyl enol ether **148**, (Scheme 46). The camphorsultam group directs the silyl enol ether to form the Z geometry **149** preferentially (Scheme 47).



Scheme 46 – Oppolzer boranesultam directed aldolization reaction.

Once the silyl enol ether has formed, the aldehyde is added, followed by a Lewis acid, for example $TiCl_4$. The Lewis acid co-ordinates to the aldehyde **146**, increasing its reactivity, which subsequently undergoes electrophilic attack *via* the Cα-*Re* face of **150**, resulting in the 2,3-anti-stereochemistry of the final product **146**.



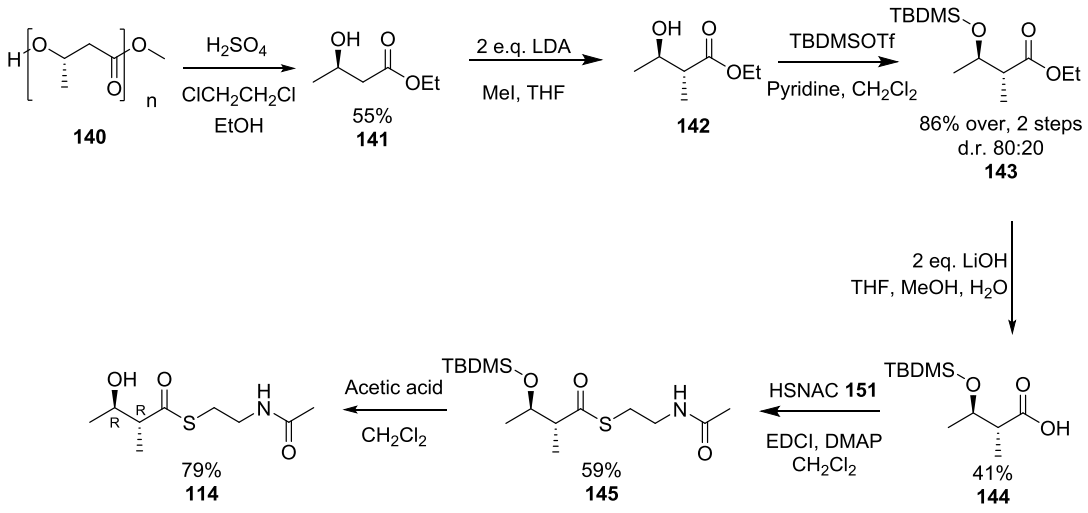
Scheme 47 – Oppolzer transition state.

2.5 Results and Discussion

2.5.1 Synthesis of diketides

2.5.1.1 Synthesis of *anti* 3-hydroxy-2-methylbutanoyl-*N*-acetylcysteamine

The synthesis of (3*R*, 2*R*)-**114** was achieved in 6 steps with an overall yield of 9% (Scheme 48).



Scheme 48 - Synthesis of (*3R*, 2*R*)-3-hydroxy-2-methylbutanoyl-*N*-acetylcysteamine.

Poly-(3*R*)-hydroxybutyrate **140** was heated in the presence of ethanol, 1,2-dichloroethane and concentrated sulfuric acid for 96 hours to give ethyl-(3*R*)-3-hydroxybutanoate **141** in a 55% yield. It was possible to carry out the synthesis of **141** on a 20 g scale.

Ester **141** was then methylated using LDA and methyl iodide at -78 °C and the hydroxyl group protected using *tert*-butyldimethylsilyl triflate (TBDMS triflate), to give (3*R*,*2R*)-**143,** in a 86% yield over the two steps. From the ^{1}H NMR spectrum the d.r. was calculated by the integration of the 3-CH_3 signals (Figure 33). A d.r. of 80:20 (4:1) was achieved in the alkylation step, the selectivity achieved in this step was lower than reported in the literature (95:5).[100] Comparison of the literature ^{1}H NMR chemical shifts for 2*R*,3*R*-**143** (2-CH_3: 1.11, d, J = 7.0 Hz; 3-CH_3: 1.16, d, J = 6.0 Hz)[111] identified the major product as 2*R*,3*R*-**143** (2-CH_3: 1.07, 3H, d, J = 7.1 Hz; 3-CH_3: 1.10, 3H, d, J = 6.2 Hz). In addition to the NMR the $[\alpha]_D$ was measured ($[\alpha]_D^{21.2}$ -36.0) and found to correlate with the value reported in the literature ([lit $[\alpha]_D$ -31.5).[116]

The acid **144** was formed by dissolving **143** in THF, water, methanol with lithium hydroxide and heating to 60 °C overnight, (2*R*,3*R*)-**144** was formed with a 41% yield. The acid **144** was reacted with *N*-acetylcysteamine **151** in the presence of DMAP and EDCI to give (*3R*,2*R*)-**145** with a 59% yield. The TBDMS protecting group was removed from **145** by stirring in the presence of acetic acid and CH_2Cl_2 for 5 days to give the product (*3R*,2*R*)-**114** with a 79% yield, in a d.r. of 80.5:19.5. It may be possible to achieve deprotection on a shorter time scale by using a stronger acid, but the conversion of starting material to product produced no side products. This allows the product **114** to be extracted with no need for further purification. After the

coupling of the acid to SNAC it was possible to purify the desired diastereomers **114**, the presence of a diastereomer was not observed in the NMR of the final compound **114** (Figure 34).

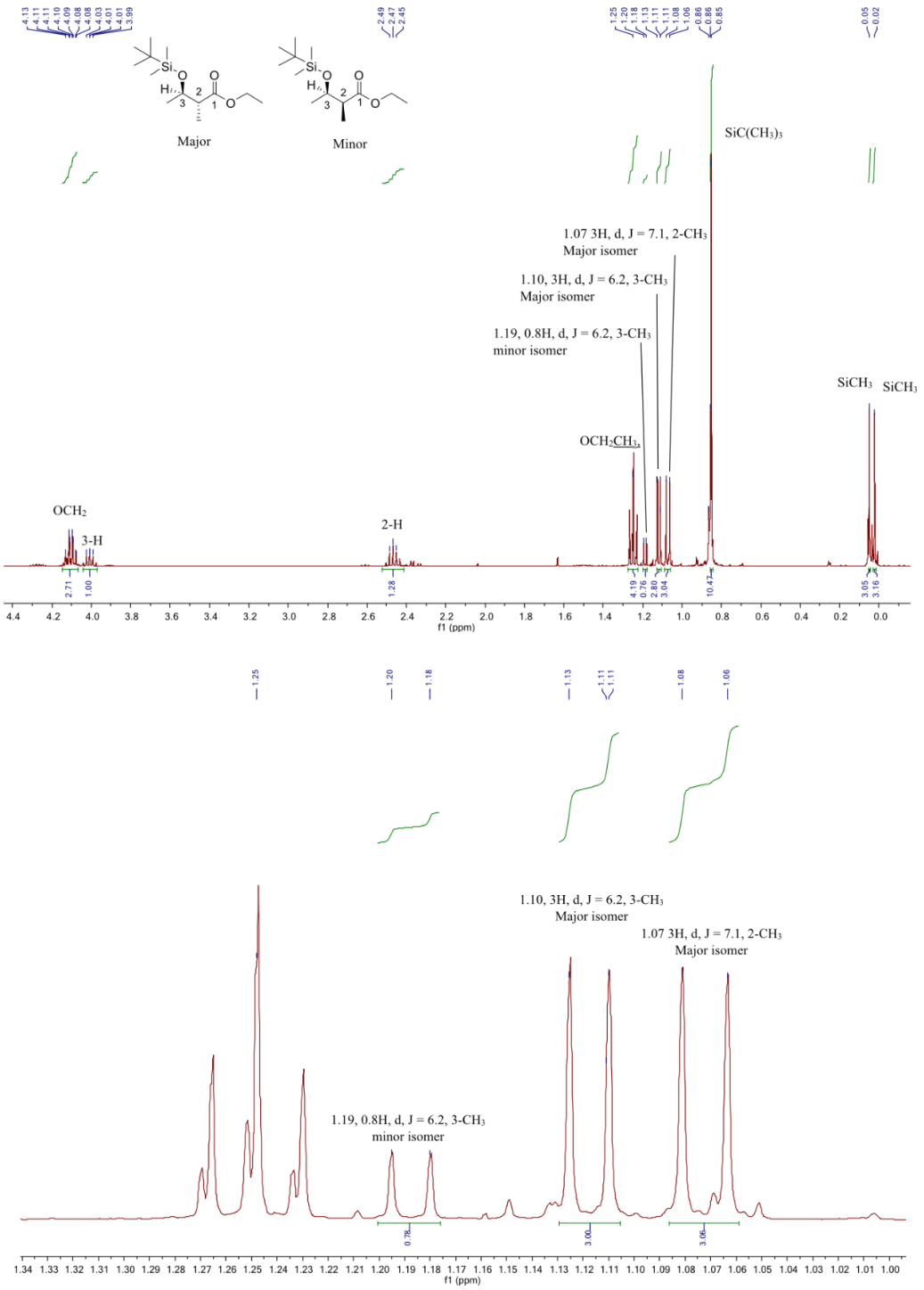


Figure 33 – ^{1}H NMR, 400 MHz, of **143** in $CDCl_3$, 3-CH_3 peak used to calculate d.r.

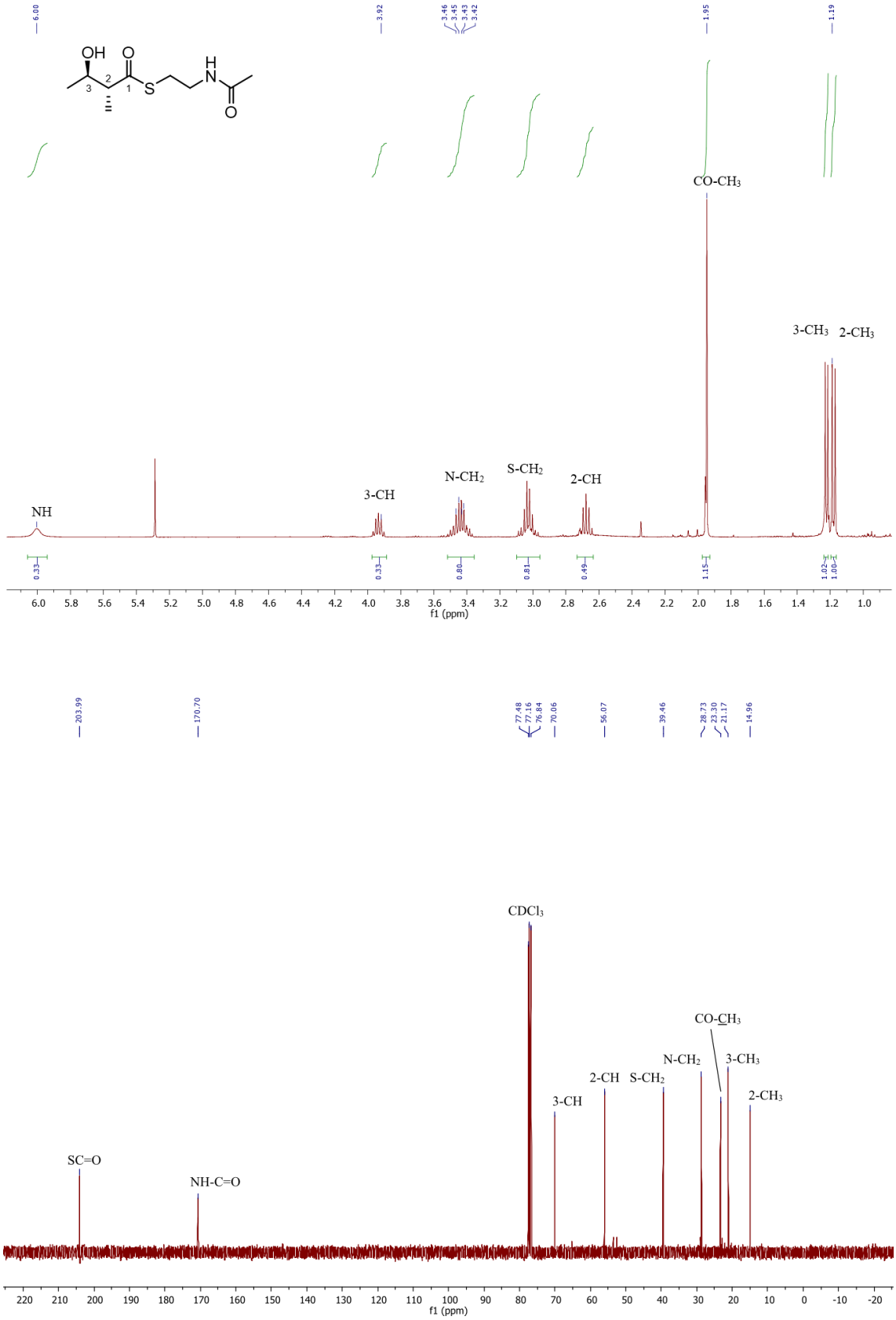
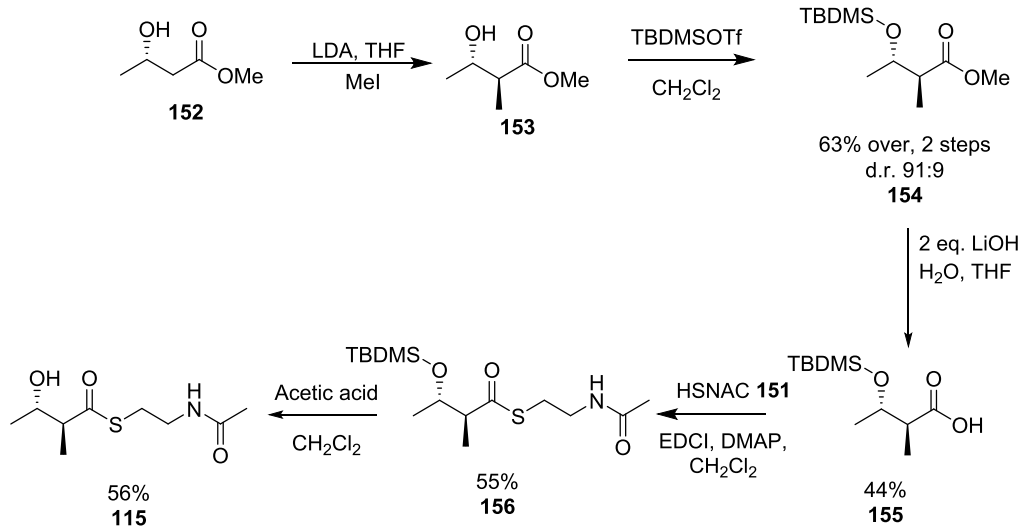


Figure 34 – ^{1}H NMR, 400 MHz and ^{13}C NMR, 100 MHz, of (3*R*, 2*R*) **114** in $CDCl_3$.

Thiolester (3*S*, 2*S*)-**115** was prepared using the same conditions used to synthesise **114** with an overall yield of 8 % from the commercially available (3*S*)-3-hydroxybutyrate **152** (Scheme 49).



Scheme 49 - Synthesis of (3*S*, 2*S*)-3-hydroxy-2-methylbutanoyl-*N*-acetylcysteamine **115**.

The methylation and protection was carried out using **152** and to give **153** resulted in an overall yield of 63%. The d.r. was measured by NMR and found to be 91:9 (10:1) in favour of the desired product (Figure 35). Two diastereomers were visible in the NMR spectrum. Comparison of the literature NMR shifts for 2*S*,3*S*-**153** (2-CH_3: 1.05, d, *J* = 7.1 Hz; 3-CH_3: 1.10, d, *J* = 6.1 Hz)[117] and the 2*R*,3*S*-diastereomer (2-CH_3: 1.13, 3H, d, *J* = 6.1 Hz; 3-CH_3: 1.14, 3H, d, *J* = 7.0 Hz)[117] identified the major product as 2*S*,3*S*-**153** (2-CH_3: 1.06, 3H, d, *J* = 7.1 Hz; 3-CH_3: 1.10, 3H, d, *J* = 6.3 Hz). The measured $[\alpha]_D$ ($[\alpha]_D^{22.5}$ +37.3) also correleated the $[\alpha]_D$ for the 2*S*,3*S*-**153** isomer ([lit $[\alpha]_D$ +37.1).[118]

Cleavage of the ester **154** afforded the acid **155** in a yield of 44%. The acid **155** was reacted with N-acetylcysteamine **151** in the presence of EDCI and DMAP to give **156** with a 55% yield. Deprotection of the hydroxyl group on **156** gave the final product **115** in a 56% yield. As was the case with **115**, over the course of the reactions the initially formed minor diastereomer was separated during purification and no longer detectable by NMR (Figure 36).

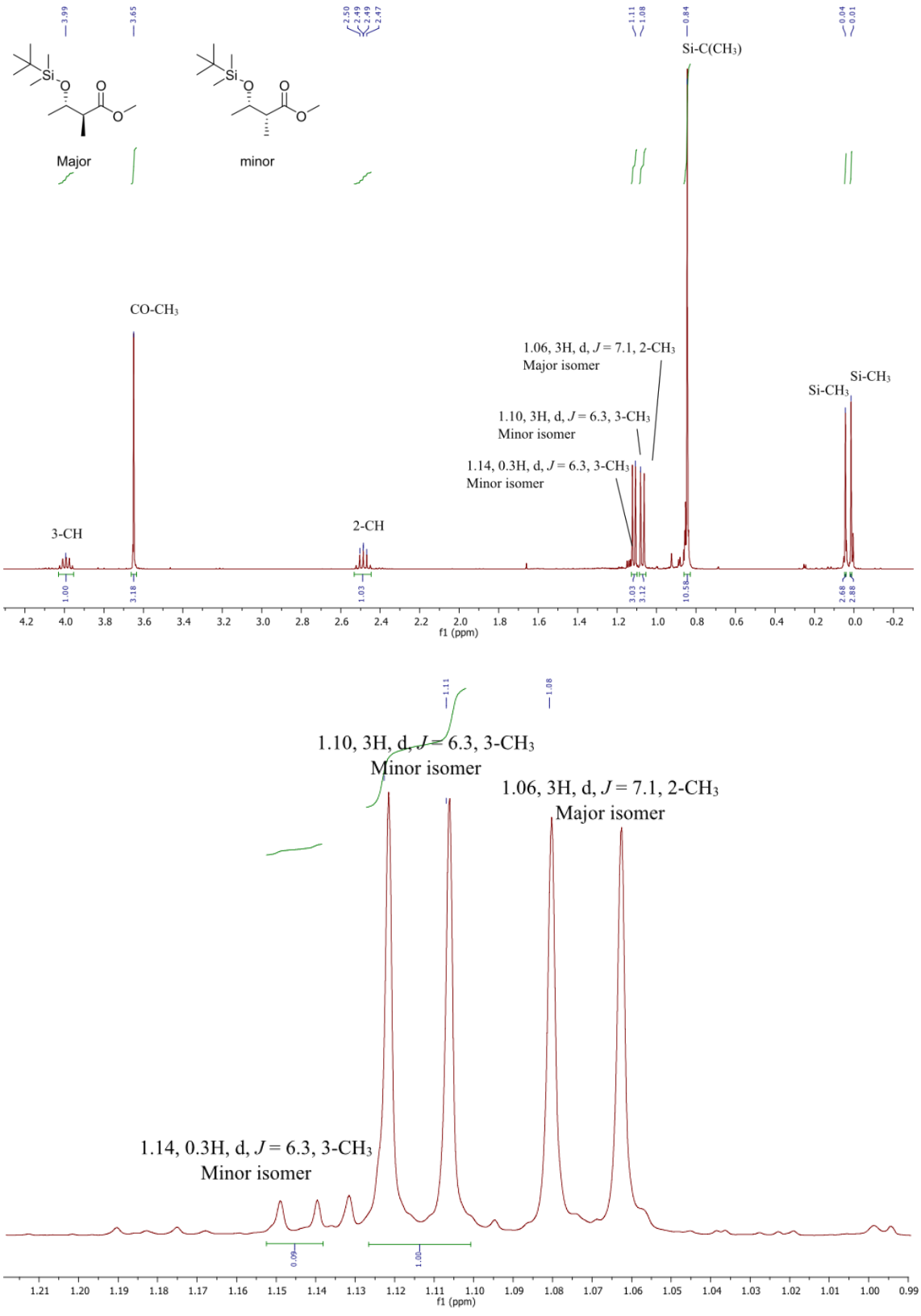


Figure 35 – ^{1}H NMR, 400 MHz, of **154** in $CDCl_3$, 2-CH_3 peak used to calculate d.r.

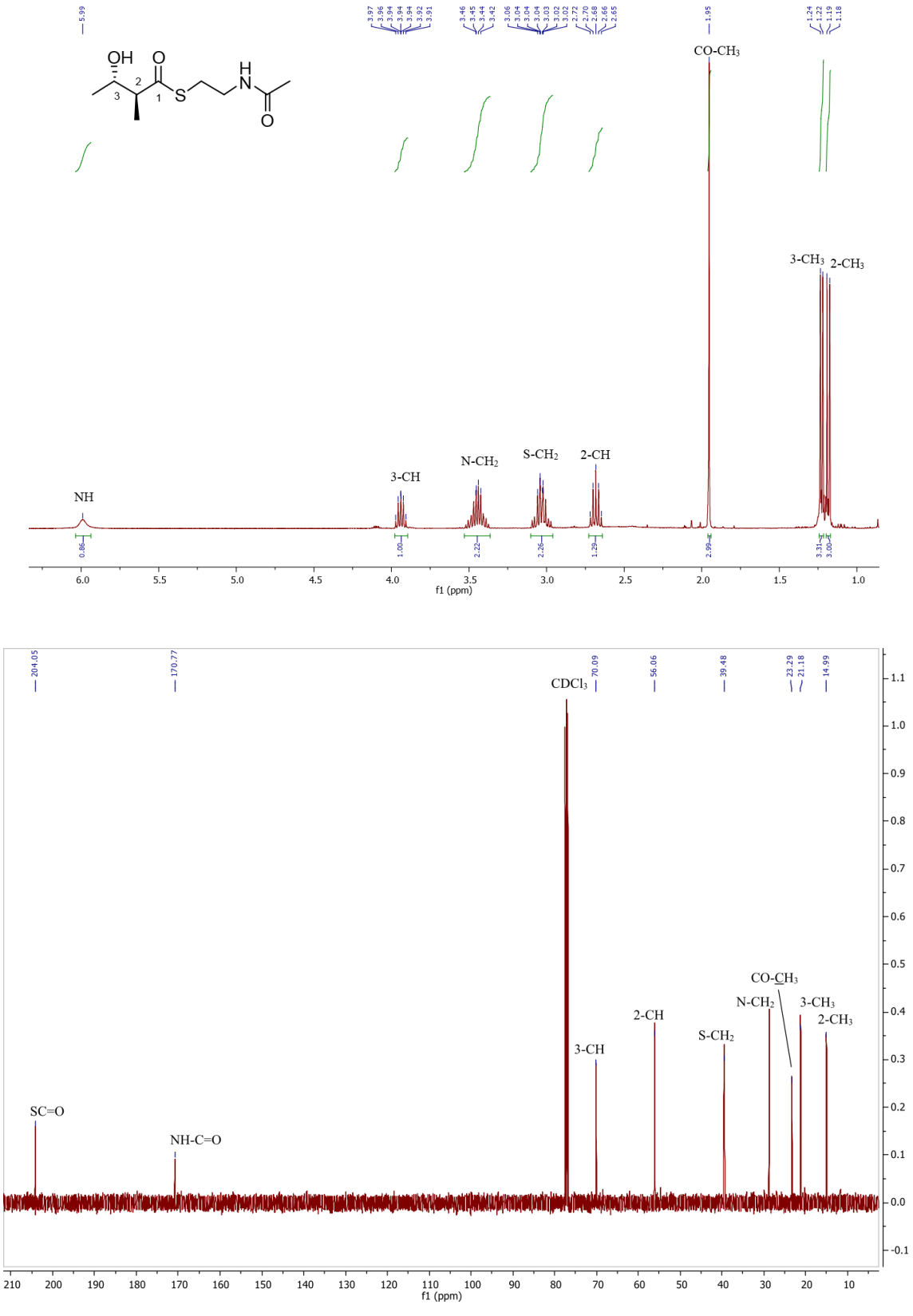
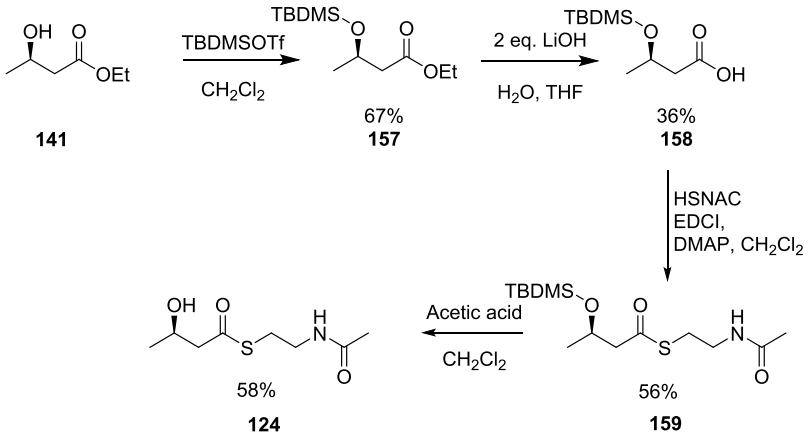


Figure 36 – ^{1}H NMR, 400 MHz and ^{13}C NMR, 100 MHz, of (3*S*, 2*S*)-**115** in $CDCl_3$.

2.5.1.2 Synthesis of (3*R*) and (3*S*)-hydroxybutanoyl-*N*-acetylcysteamine

The synthesis of substrate mimics (3*R*)-**124** and (3*S*)-**125** was achieved using the experimental conditions used previously in the synthesis of **114** and **115**. The synthesis of (3*R*)-**124** was achieved in five steps with an overall yield of 4% The yields of each reaction are summarised in Scheme 50 and the ^{1}H NMR of **124** in Figure 37. The enantiomer (3*S*)-**125** was prepared using identical conditions to **124** from the commercially available (3*S*)-3-hydroxybutyrate **152** and the synthesis was achieved in four steps with an overall yield of 8%.



Scheme 50 - Synthesis of (3*R*)-**124**.

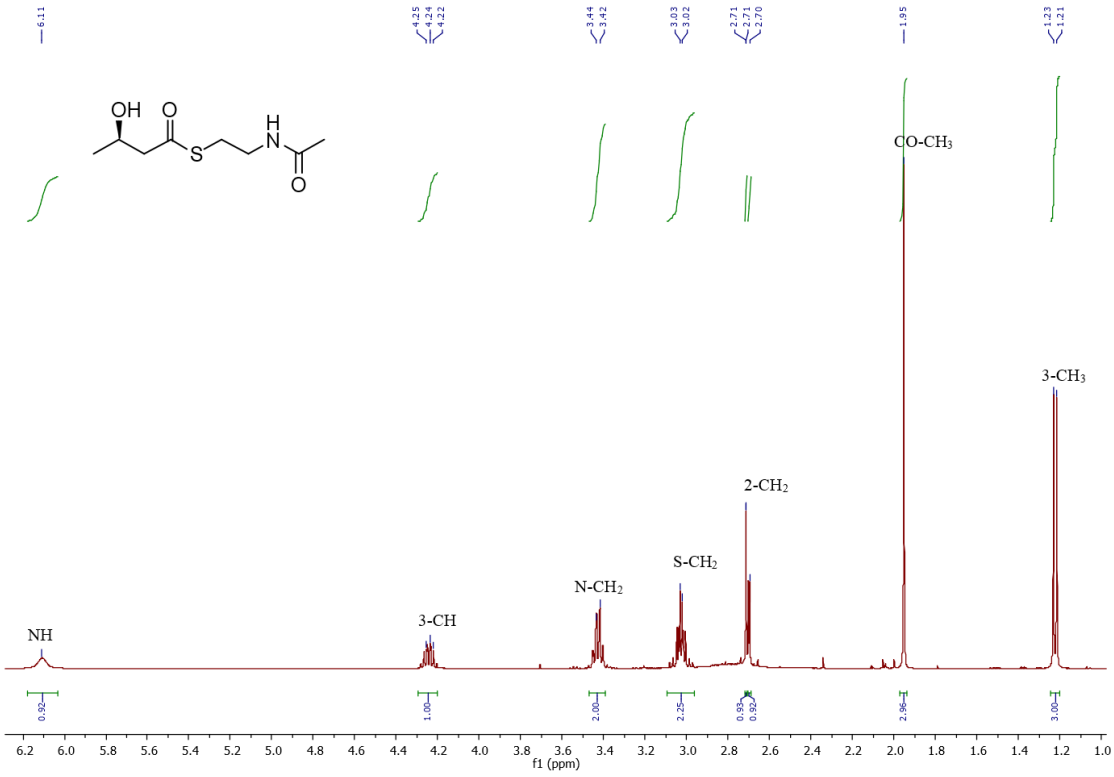
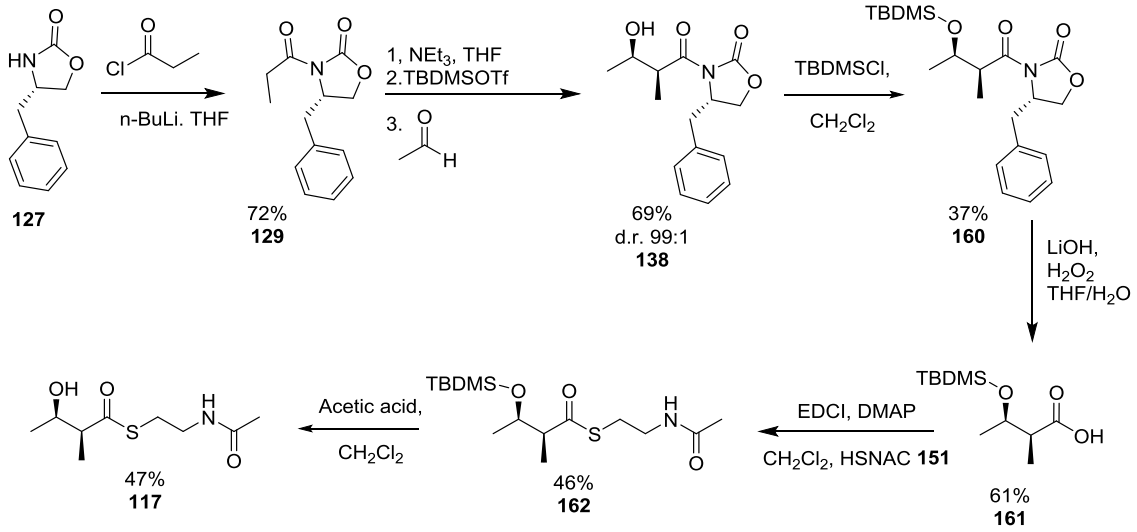


Figure 37 - ^{1}H NMR, 400 MHz, of (3*R*)-**124** in $CDCl_3$

2.5.1.3 Synthesis of *syn* 3-hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 117

The synthesis of (2*S*,3*R*)-**117** and (2*R*,3*S*)-**116** was carried out using oxazolidinones auxiliaries **129** and **130** in 6 steps, with overall yields of 2% and 9% respectively. The summary of (2*S*,3*R*)-**117** reaction yields is summarised in (Scheme 51). (*S*)-4-Benzyl-2-oxazolidinone **127** was reacted with propionyl chloride to give the acylated Evans auxiliary **129** as a white solid in a 72% yield. The acylation of (*R*)-4-benzoyl-2-oxazolidinone gave **130** in a 99% yield.



Scheme 51 - Synthesis (2*S*,3*R*)-**117**.

The second stage of the synthesis, the formation of the aldol **138** was problematic. The acylated Evans auxiliary **129** was reacted with acetaldehyde in the presence of dibutyl boron triflate and triethylamine (Scheme 51). Initial reactions carried out using the procedure described by Harris *et al*.[102] were unsuccessful, with either very low or no formation of the desired product.

Literature research identified that the purity and dryness of reagents was vital for a successful, high yielding reaction.[105, 119] Triethylamine (NEt_3) was dried over potassium hydroxide prior to use and stored under nitrogen in the presence of molecular sieves to ensure adequate dryness on use. Acetaldehyde was distilled from calcium hydride before every reaction. The purification of the commercially purchased boron triflate was deemed too hazardous due to the reagent's exothermic nature.

NMR analysis of the dibutyl boron triflate reagent was undertaken with the ^{1}H NMR (Figure 38, A.) showing the one triple peak, indicative of a CH_3 group at the end of the butyl chains. Two other proton environments were present, accounting for the three CH_2 groups of the butyl chains. ^{19}F NMR (Figure 38, B.) showed only one fluorine environment was present, and one environment was also detected in the ^{11}B NMR (Figure 38, C.). If the product had degraded, a more complicated NMR spectra would have been expected. The reagent was deemed fit for purpose and undegraded and was therefore used without further drying or purification.

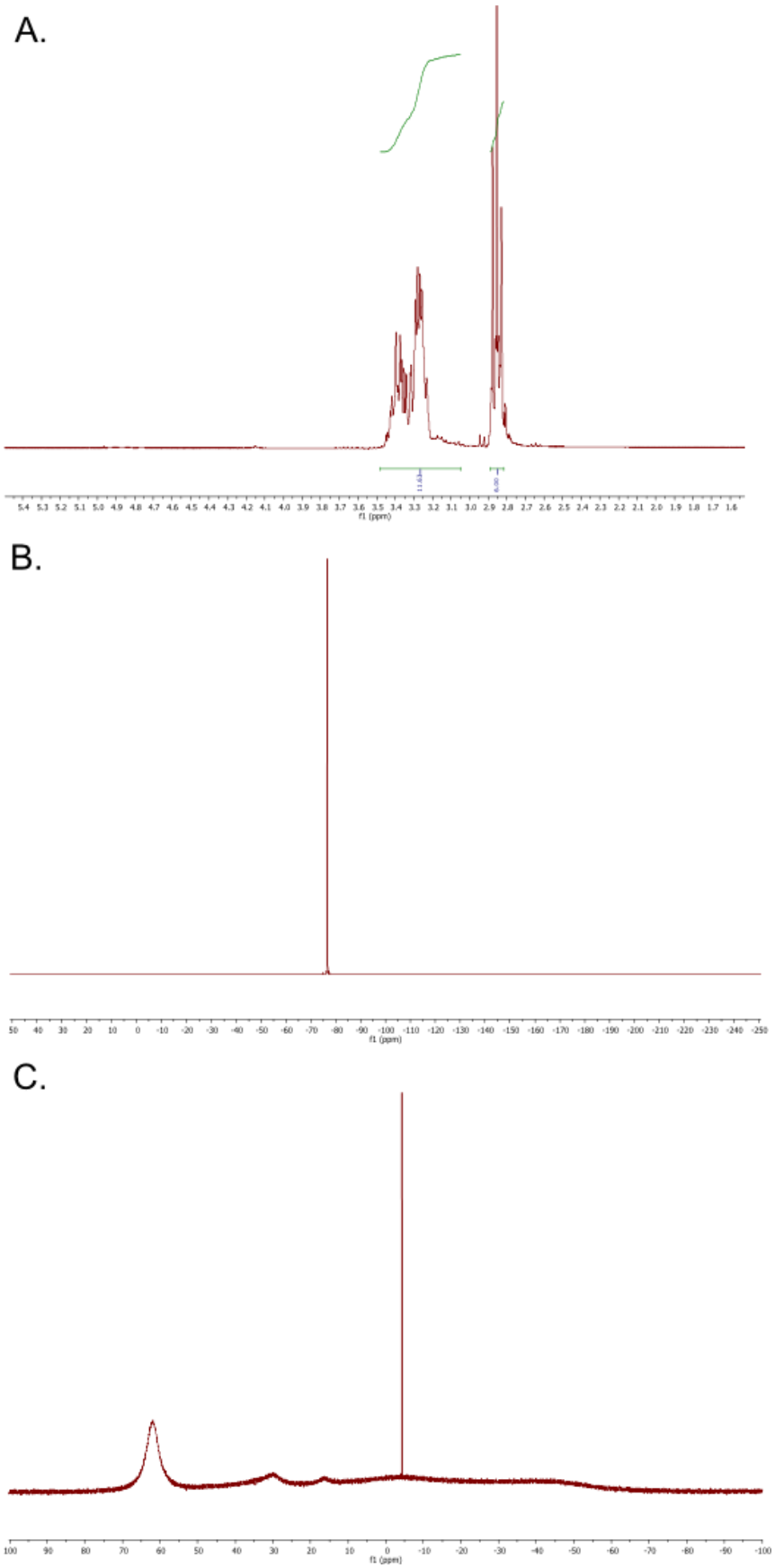
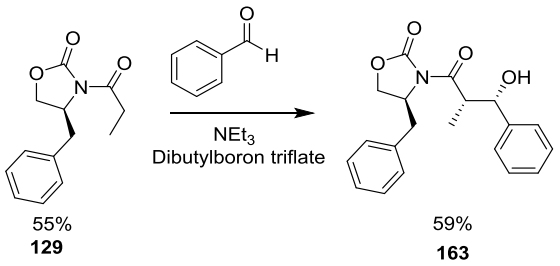


Figure 38 – **A**, ^{1}H NMR, 400 MHz, of dibutyl boron triflate reagent; **B,** ^{19}F NMR, 300 MHz of dibutyl boron reagent; **C,** ^{11}B NMR, 300 MHz of dibutyl boron triflate solution.

A reaction was carried out using a different aldehyde (Scheme 52). Benzaldehyde was chosen, as it possessed an aromatic ring, making it less volatile to handle and monitor the progress of the reaction via TLC. Benzaldehyde and NEt_3 were distilled over calcium hydride and stored under an inert nitrogen atmosphere prior to use. The test reaction with benzaldehyde was successful and formed **163** but a subsequent reaction with acetaldehyde was poor yielding and with low stereo selectivity.



Scheme 52 - Reaction scheme of Evans auxiliary and benzaldehyde.

The Evans aldol reaction proceeds *via* a pseudo chair closed transition state (Scheme 40) and many variables need to be correct for the desired product to be formed. There are three variables that could result in the incorrect stereochemical outcome. The first is the formation of the enolate; it is possible that the temporary chiral enolate could take the *E* rather than the *Z* geometry, this would affect the α position.

Secondly the aldehyde could approach the wrong face of the enolate, affecting the stereochemistry of the α and β positions. Finally, the aldehyde could orientate with the hydrogen in the equatorial rather than in the axial orientation. This would result in the wrong stereochemistry at the β position. It is most likely that the aldehyde would take the wrong orientation.

When the aldol reaction was carried out using benzaldehyde only one stereoisomer was formed. Comparison of the literature NMR and the $[\alpha]_D$ for *syn* and *anti* products confirmed the formation of the *syn* product (experimental: $[\alpha]_D^{22}$ +47.5 (c 1.0, $CHCl_3$), *syn* lit: $[\alpha]_D$ 70.4 (c 1.04, $CHCl_3$)[102], *anti* lit: $[\alpha]_D$ -8.6 (c 1.51, $CHCl_3$)[120]). The benzaldehyde is a bulky reagent, so the steric clash from the desired orientation between the aldehyde and the Evans benzyl group would be highly unfavourable. The acetaldehyde on the other hand, is a small methyl group. The steric repulsion between the aldehyde and the Evans benzyl group would be reduced, resulting in a greater likelihood of the wrong orientation occurring and a reaction taking place. It was found that the length of time the reaction was stirred before the addition of acetaldehyde was vital for high selectivity. Stirring the reaction mixture at -78 °C for 1 hour and then 100 minutes at 0 °C before the addition of acetaldehyde, allowed the desired *Z*-enolate to form, resulting in an increased selectivity.

The Evans derivatives **129** and **130** were reacted with acetaldehyde in the presence of dibutyl boron triflate and trimethylamine to form the aldol (2*S*, 3*R*)-**138** and (2*R*, 3*S*)-**164** respectively. The oxazolidinone-directed aldol reactions resulted in high stereoselectivity. NMR was used to measure the d.r. of each compound, (2*S*, 3*R*)-**138** was synthesised with a d.r. of 100:1 and (2*R*, 3*S*)-**164** was synthesised with a d.r. of 100:3. The selectivity of these synthesise are comparable to those quoted in the literature by Evans *et al.* and Hayashi *et al.*[105, 107] The $[\alpha]_D$ of the two aldol products were measured (2*S*, 3*R*)-**138**: $[\alpha]_D^{22}$ +48.8 (c 1.0, $CHCl_3$) and (2*R*,3*S*)-**164** $[\alpha]_D^{22}$ -47.2 (c 1.0, $CHCl_3$). Comparison of the measured $[\alpha]_D$ to the literature $[\alpha]_D$ values of the *syn* and *anti* products confirmed the formation of the *syn* aldol product (*syn* Lit. (2*S*,3*R*)-**129**:$[\alpha]_D$ +51.1 (c 1.0, $CHCl_3$),[102, 121] (2*R*,3*S*)-**164**: $[\alpha]_D$ -58.8 (c 1.0, $CHCl_3$)[122] and *anti* (2*S*,3*S*): $[\alpha]_D$ -19.8 (c 0.5, $CHCl_3$).[122]

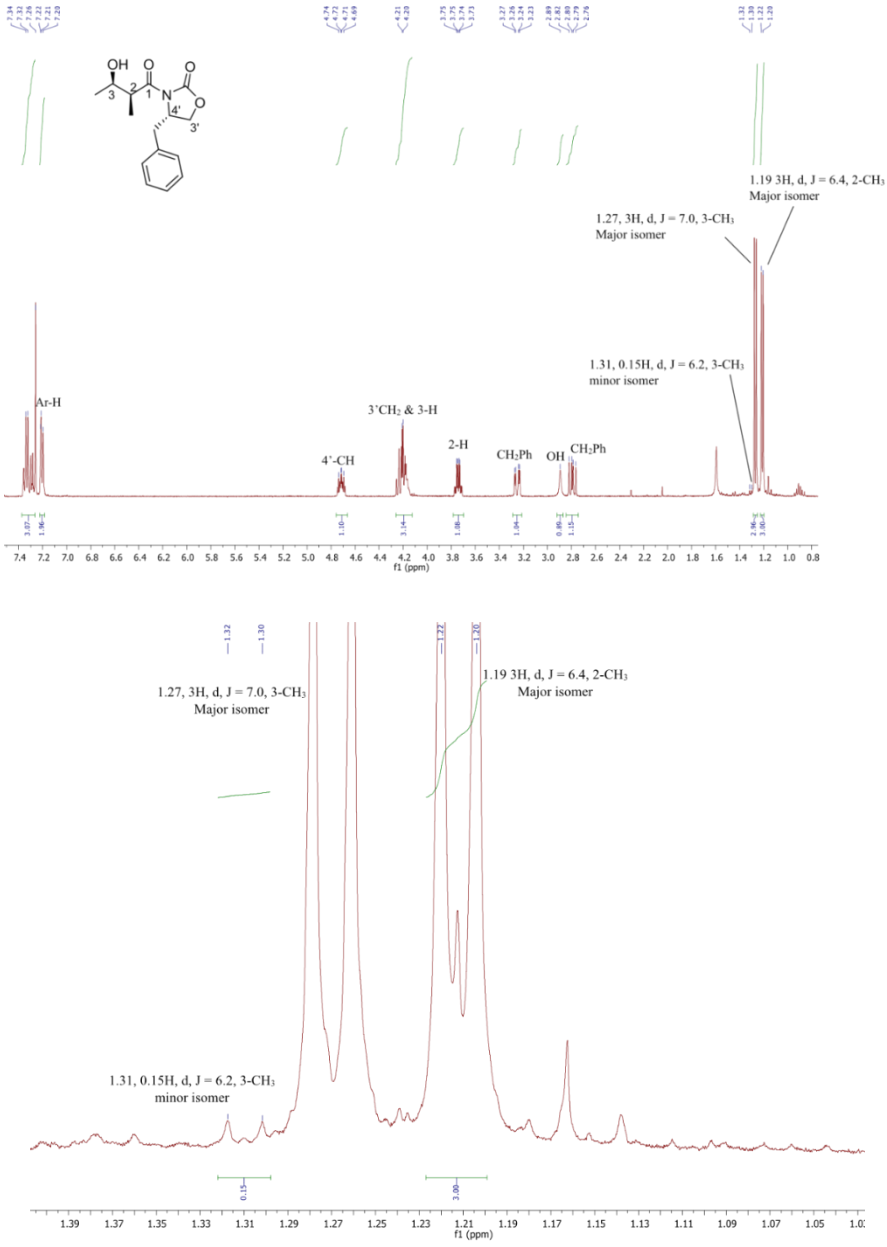


Figure 39 - ^{1}H NMR, 400 MHz, of (2*S*, 3*R*)-**138** in $CDCl_3$, 2-CH_3 peak used to calculate d.r.

After the formation of the aldols (2*S*,3*R*)-**138** and (2*R*,3*S*)-**164**, they were subsequently treated with *tert*-butyldimethylsilyl chloride to give the silyl ethers (2*S*,3*R*)-**160** and (2*R*,3*S*)-**165**. The Evans auxiliary was cleaved in the presence of hydrogen peroxide and lithium hydroxide to give the acids (2*S*,3*R*)-**161** and (2*R*,3*S*)-**166**. Acids (2*S*,3*R*)-**161** and (2*R*,3*S*)-**166** were treated with *N*-acetylcysteamine in the presence of EDCI and DMAP to form (2*S*,3*R*)-**162** and (2*R*,3*S*)-**167**. Removal of the protecting group was achieved by stirring (S*R*,3*R*)-**162** and (2*R*,3*S*)-**167** in the presence of acetic acid for 5 days giving the final compounds (2*S*,3*R*)-**117** (Figure 40) and (2*R*,3*S*)-**116**. No diastereomer was observable in the NMR of either compound.

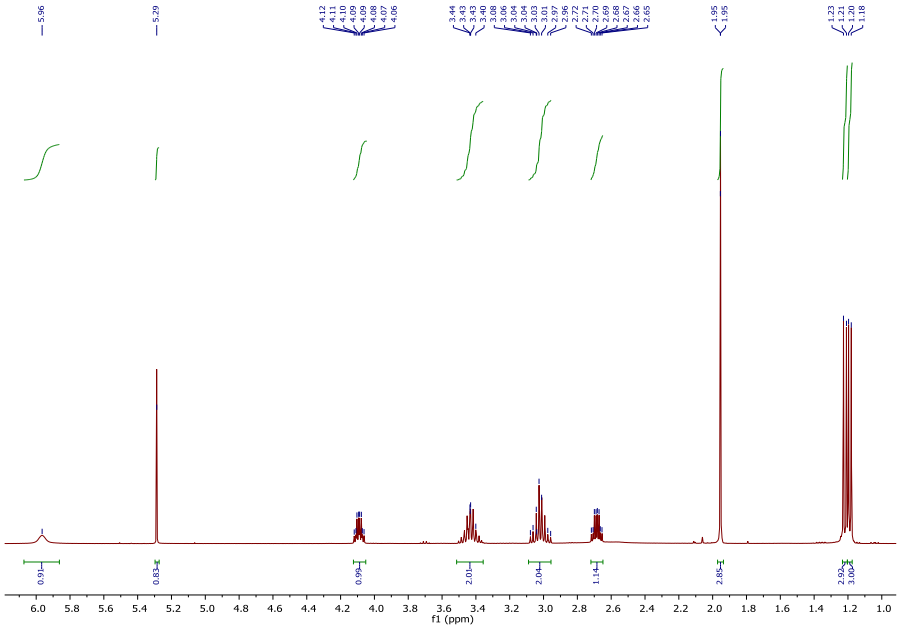
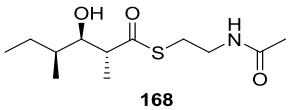


Figure 40 - ^{1}H NMR, 400 MHz, of (2*S*,3*R*)-**117** in $CDCl_3$.

2.5.2 Synthesis of triketide substrate mimic

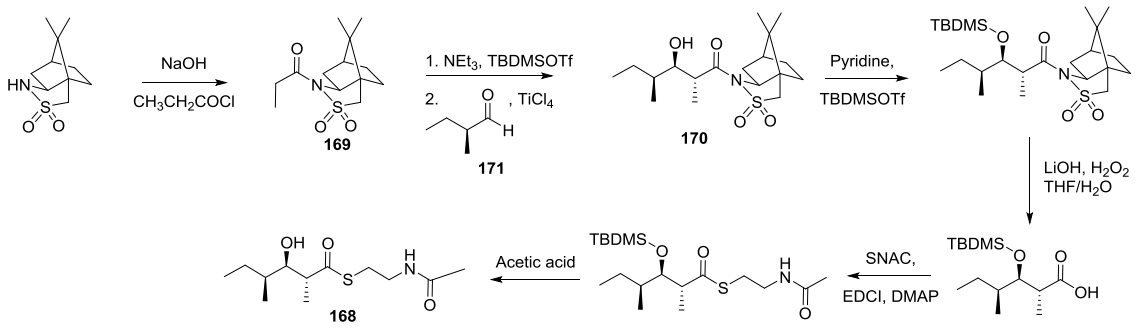
The results from the selectivity assays (discussed in chapter 4) show the SQTKS DH domain to be highly selective. It is probable that the DH would show the same selectivity towards triketides as was observed towards diketide substrate mimics, therefore only the triketide **168** would be synthesised in order to test this theory.



168

Figure 41 – triketide substrate mimic.

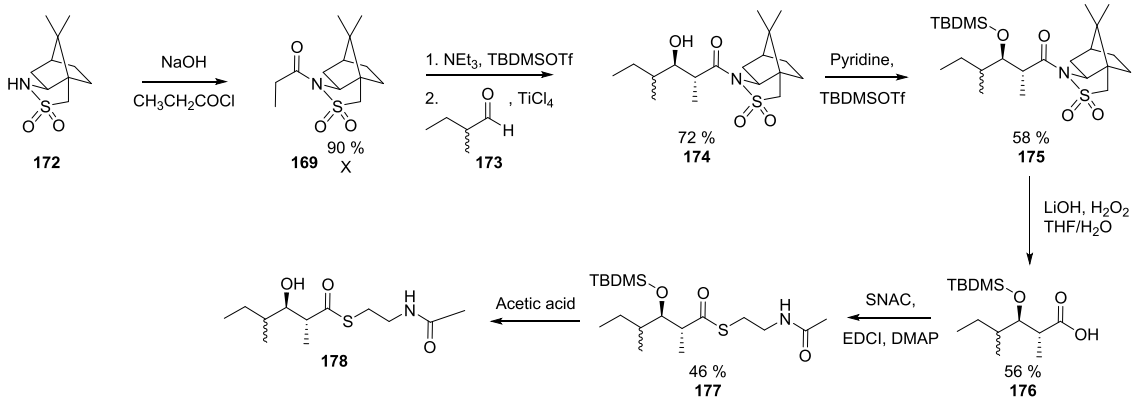
Utilizing the camphorsultam **169** directed aldol reaction, synthesis of **168** was proposed, (Scheme 53).



Scheme 53 – Proposed synthesis for **168**.

2.5.2.1 (4*RS*,3R,*2R*)-2,4-dimethyl-3-hydroxyhexanoyl-*N*-acetylcysteamine 178

Initially the aldol reaction to form **170** was to be carried out using (2*S*)-methylbutyraldehyde **171**, unfortunately the compound was not available commercially, but it could be prepared by oxidising (*S*)-(-)-2-methylbutan-1-ol with TEMPO in the presence of KBr in water.[123] However, isolation of the product in a dry enough state to allow the aldol reaction to occur proved difficult due to the volatile nature of the product. The aldol reaction was attempted using crude (2*S*)-methylbutyraldehyde **171** in DCM, rather than isolating the product fully. Attempts were made to dry the (2*S*)-methylbutyraldehyde **171**/DCM solution prior to aldol reaction with $MgSO_4$, but none of the required product **170** formed. 3 and 4 Å molecular sieves were also added prior to the aldol reaction, but again this led to none of the required product. In the interest of time, it was decided to carry out the aldol reaction step using racemic ±2-methylbutanal **173** (Scheme 54).



Scheme 54 – Synthesis of (3*R*,*2R*)-2,4-dimethyl-3-hydroxyhexanoyl-*N*-acetylcysteamine (2*R*, 3*S*)-**178**.

(1*R*)-(+)-Camphorsultam **172** was reacted with propionyl chloride to give acylated sultam **169** in a 90% yield. Acylated sultam **169** underwent an aldol reaction with ±2-methylbutanal **173** in the presence of triethylamine and titanium (IV) chloride to form **174** in 72% yield. The aldol product **174** was subsequently reacted with *tert*-butyldimethylsilyl triflate to protect the hydroxyl group and gave **175** in a 58% yield. The camphorsultam group was cleaved by stirring **175** with hydrogen peroxide and lithium hydroxide in solvent to give the acid **176** in 56% yield. Acid **176** was reacted with *N*-acetylcysteamine in the presence to EDCI and DMAP to form **177**

in 46% yield. The removal of the TBDMS protecting group on **177** to give **178** was attempted using the gentle method used on diketide SNAC substrate previously, by stirring at room temperature in acetic acid. After 4 days, only a small amount of deprotected product was observed. Deprotection using p-toluenesulfonic acid (pTsOH) was also attempted, but was unsuccessful. Further methods of deprotection could not be attempted due to time constraints.

2.6 Conclusion

Six of the natural and non-natural diketide *N*-acetylcysteamine substrates were synthesised. The natural methylated SQTKS DH *anti*-substrate mimics, **114** and **115**, were accessed using Fráter-Seebach alkylation. The selectivity of the initial alkylation step to form (*3R*,2*R*)-**142** and (*3S*,2*S*)-**153** was found to be lower than reported in the literature (95:5) [100] with a selectivity of 77:23 and 91:9 respectively. Although the selectivity of the alkylation step was lower than desired, over the course of the subsequent reactions, it was possible to separate the two diastereomers. The NMR of the final compounds showed only one diastereomer was present (2*R*,3*R*)-**114** (Figure 34) and (2*S*,3*S*)-**115** (Figure 36) and used in the assays discussed in chapter 4.

The synthesis of the *syn* products, (2*R*,3*S*)-**116** and (2*S*,3*R*)-**117** using Evans auxiliary directed aldol reactions gave a greater selectivity towards the desired diastereomers of 99:1 and 97:3 for (2*S*,*3R*)-**138** and (2*R*,*3S*)-**164** respectively comparable to the literature d.r. of 98:2.[107] Subsequent hydroxyl group protection and purification of compounds (2*S*,*3R*)-**160** and (2*R*,*3S*)-**165** allowed the unwanted minor diastereomer to be removed. There was no observable diastereomer visible in the NMR of the final compounds (2*S*,*3R*)-**117** (Figure 40) and (2*R*,*3S*)-**116** . Overall the yields for each stage of synthesis were acceptable, between 60 and 80%. The formation of the acids (*2R*,*3R*)-**144**, (*2S*,*3S*)-**155**, (*3S*)-**179** and (*3R*)-**158** were all lower than desired, with yields ranging from 36-44%.

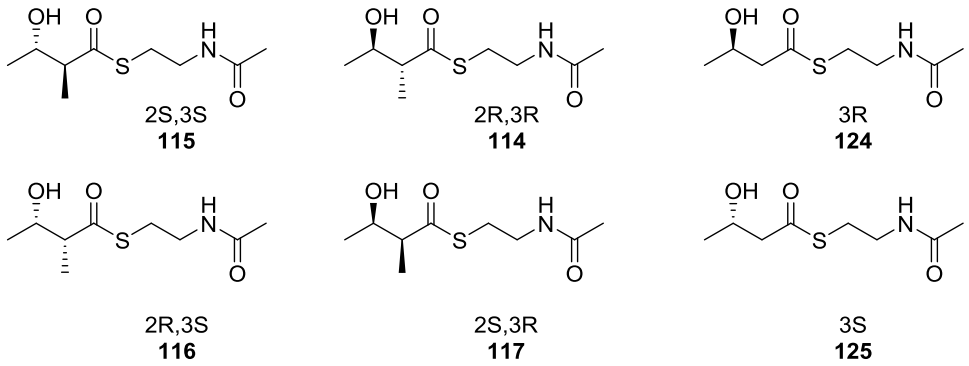


Figure 42 – synthesised diketic mimics.

Synthesis of the triketide substrate mimic **168** proved more difficult. Initially the camphorsultam aldol reaction was attempted using (2*S*)-methylbutanal **171** but was unsuccessful. When the camphorsultam directed aldol reaction was carried out using racemic ±2-methylbutanal **173**, the reaction was successful. Cleavage of the auxiliary and coupling to the SNAC was achieved with no issues. The deprotection of **177** to afford the final compound **178** proved problematic and none of the target compound **178** was isolated.

2.7 Future work

The diketide substrates were synthesised as *N*-acetylcysteamines. While this was suitable for assessing the selectivity and kinetics of the DH domain, discussed in detail in Chapter 4. Synthesis (2*R*,3*R*)-**114** as a pantetheine derivative may allow a more accurate assessment of the K_M of the DH domain. Further investigation into the correct conditions for the deprotection of the triketide substrate **177** to give **178** is also required.

Chapter 3 - Optimisation of DH protein expression

3.1 Introduction

Although there are many systems available by which it is possible to heterologously express protein, the bacterial system *Escherichia coli. (E. coli)* remains one of the most popular.[124] *E. coli* is attractive as it is easy to genetically manipulate, inexpensive to culture and has the ability to undergo rapid growth.[125] While there are many advantages to using *E. coli* to produce protein, there are also a number of disadvantages. The rapid growth allows proteins to be produced in one day, but the rapid production can result in unfolded, misfolded or partially folded insoluble protein due to the speed at which *E. coli* is able to transcribe and translate the target genes. A wide number of mutant host strains, cloning vectors and strategies have been developed to help overcome these difficulties. There are many parameters which can be varied in order to optimise an expression protocol, these include cell strain, media composition and growth conditions.

In the initial stage of protein production, the genes of interest are cloned into a plasmid containing an antibiotic resistance gene and T7 RNA promoter.[126] The T7 promoter contains a lac operon which allows expression of the desired protein to be controlled by the addition of lactose or Isopropyl β-D-1-thiogalactopyranoside (IPTG). The second stage involves the transformation of the plasmid into the desired *E. coli* strain. The presence of an antibiotic resistance gene allows selection of correctly transformed *E. coli*.. Routinely *E. coli* BL21 and K12 or their derivatives are used for protein expression.[124] In the final stage, the transformed *E. coli* is grown and the protein expression induced, before harvesting the cells and extracting the protein.

There are many examples in literature, of isolated DH domains which have been produced as soluble active proteins using a range of cell strains and growth conditions. Borrelidin DH2 was produced using *E. coli* BL21 (DE3)-CodonPlus RP cells, grown in 2TY media, with the expression of the protein being controlled by IPTG.[60] The DH domains from Pikromycin DH, (PikDH2),[67] Curacin DH domains (CurF, CurJ, CurK, CurH)[17] rifamycin, (RifDH10)[13] and erythromycin (EryDH4)[6] have all been expressed using the *E. coli* BL21 (DE3) cell strain. PikDH10 and the curacin DH domains used TB media,[17, 67] while RifDH10 and ErDH4 were cultured in LB media.[6, 13] Protein expression in all of these examples was induced by various concentrations of IPTG.

3.2 Previous work

Previous work, undertaken by David Ivison, investigated the possibility of isolating the individual catalytic domains from SQTKS and producing them as isolated proteins.[95] SQTKS was systematically dissected to give constructs of isolated catalytic domains such KR and DH or groups of domains such as KR-ACP or KS-AT-DH. In total 22 plasmids containing these

constructs were produced and subjected to high throughput screening to identify which domains expressed as soluble protein. One of the isolated domains to be expressed as soluble protein in reasonable amounts, was construct #7615, SQTKS DH.[95]

Expression of the Construct #7615 was scaled up in order to isolate enough protein to carry out assays to assess the isolated domain's activity. The construct was transformed into *E. coli* BL21(DE3) and grown on 2TY media. Following IPTG induction at 16 °C overnight, cells were harvested and lysed by sonication. The SQTKS DH was isolated and purified. Once pure, the protein was incubated with racemic 2-methyl-3-hydroxy-SNAC **111**, and the domain was found to be biochemically active. Full details of this assay and subsequent assay work is detailed in chapter 4. To explore the programming of the SQTKS DH domain it was vital that sufficient quantities of soluble active protein could be produced. Although the SQTKS DH had previously been produced, there was much scope in the protocol for optimisation.

3.3 Results

3.3.1 Optimisation of expression strain and media

When over expressing protein in *E. coli* there are two commonly encountered problems: poor expression and insoluble protein.[124, 127] The work investigating the DH domain encountered many of these problems, this set of experiments investigates both strain and medium.

A number of bacterial host strains have been developed to help overcome many of the issues encountered when heterologously expressing protein.[125] The *E. coli* BL21 strain and its many derivatives are routinely used to express protein on an industrial scale.[125, 127] This strain and its derivatives are *lon* and *ompT* protease deficient, this genetic mutation gives increased protein stability.[127] *Lon* protease, also known as protease *La*, is a homo-oligomeric ATP-dependent protease.[128] Its main function is to act as a quality control protease, degrading misfolded or recombinant proteins.[129] *OmpT* is also a housekeeping protease, which degrades endogenous or fusion proteins which come into contact with the outer membrane.[130] Work by Gottesman *et al*. showed deletion of *lon* and *ompT* resulted in increased expression and stability of recombinant proteins.[125]

A common reason heterologous expression fails is the presence of codons in the target's mRNA which are used infrequently by *E. coli*. The presence of rare codons such as arginine (AGA, AGG) and isoleucine (AUA) can cause translational stalling which can result in low expression of proteins or expression of truncated protein. Two approaches can be taken to overcome this problem. Firstly, it is possible to quantify the number and the locations of rare codons using websites such as RaCC (the rare codon calculator; http://nihserver.mbi.ucla.edu/RACC). This allows genes to be "codon optimized" where rare codons are replaced with codons more commonly found in the host, either by site directed mutagenesis or gene synthesis. The second

approach is to modify the expression host. Alternative *E. coli*. strains are available which can co-express genes encoding for rare tRNAs, for example Rosetta cells from Novagen.

An initial cell screen was carried out by Dr Alan Scott to identify a suitable strain. The cell screen showed that both the BL21 (DE3) and Rosetta (DE3) strains produced the DH protein. The level of expression in each strain was comparable. The expression construct contained the native fungal sequence and had not been codon optimised. It was decided to proceed using the *E. coli*. Rosetta (DE3), as this strain encodes a number of rare codon tRNAs.[125] Further work with the Rosetta (DE3) strain encountered a number of issues; the two main problems were inconsistent growth of cultures and low or inconsistent expression of soluble protein.

It was thought that the inconsistent growth of cultures could be due to several reasons. Loss of the plasmid from the transformed *E. coli* cell would result in loss of antibiotic resistance and subsequent cell death. The protein produced could be toxic to the *E. coli* causing cell death. Although the plasmid has been designed to only express the protein under certain induced conditions, some background expression may occur. Background expression of the protein would result in less than optimal growth of the *E. coli* if the protein were toxic, as cell death would occur during the growth phase of the culturing.

In order to try and overcome these problems a number of expression trials were undertaken. The effect of cell strain, media, growth conditions and antibiotic loading on protein production were investigated (Figure 43).

The following cell strains were trialled: Rosetta, JM109, BL21 (DE4), C43(DE3) and BL21 (DE3) CodonPlus – RP. C43(DE3), a mutant strain of *E. coli* BL21(DE3), was chosen as it has frequently been used to overcome toxicity associated with over expression of proteins using a T7 RNA polymerase expression system.[131] BL21 (DE3) CodonPlus - RP strain was chosen as it is optimised for expression of proteins which contain rare codons, not normally found in *E. coli*. Unlike the Rosetta strain, which is codon optimised for arginine (AGG/AGA, CGS), isoleucine (AUA), leucine (CUA), proline (CCC) and glycine (GGA), BL21(DE3) CodonPlus – RP is optimised for arginine (AGG, AGA) and proline (CCC).

The standard expression protocol, detailed in Section 7.37.3, was followed using IPTG (final concentration, 0.4 mM) to induce expression. Non-induced conditions were also trialled, the standard expression protocol was followed and incubated at 37 °C with shaking until the OD_{600} reached 0.6. Once the appropriate OD_{600} had been reached, the temperature of the culture was then adjusted to the expression temperature of 19 °C and no induction agent added.

The investigation trialled Rosetta in rich medium (enriched with metal salts and other additives), as it was thought it would help improve the growth of the *E. coli*, allowing a higher cell density to be reached, leading to an increased yield of protein. All other cell strains were

grown in 2TY media under the following conditions: standard conditions, low antibiotic loading, additional antibiotic loading and media exchange.

As toxicity of over expressed protein is a common problem it was decided to trial media exchange as a way of overcoming this problem. Media exchange conditions consisted of the production culture being grown at 37 °C to an $O.D_{600}$ of 0.6. The cells were then harvested gently and suspended in fresh production media. Expression of the culture was then induced with IPTG and the incubation temperature reduced to 19 °C. The exchange of media before induction allows the removal of any toxins which may have built up in the media during the initial growth of the cells. Glucose was added to the media to reduce the level of background protein expression as it prevents the activation of the T7 promotor.

In order to overcome the potential loss of the plasmid from the transformed cells and therefore subsequent loss of antibiotic resistance, growth conditions involving high loading of carbenicillin (0.5 mM) and low loadings of carbenicillin (0.2 mM) were trialled.[132]

In addition, some flasks were topped up with additional antibiotic. Additional antibiotic conditions consist of the production culture being grown at 37 °C to an $O.D_{600}$ of 0.6. Expression of the culture is then induced with IPTG and the incubation temperature reduced to 19 °C. Upon induction an additional 200 µl of 0.2 M carbenicillin stock was added to 200 mL flasks of low antibiotic loading, final carbenicillin concentration was 0.4 mM. 500 µl of 0.2 carbenicillin stock was added to 200 mL flasks of high antibiotic loading final carbenicillin concentration was 1.0 mM.

The range of cell strains selected were transformed with the plasmid #7615. All strains were transformed successfully and grew well on LB Glu/Carbenicillin plates apart from BL21 (DE3), where no colonies were observed. Single colonies from each successful transformation were used to inoculate 50 mL LB Starter cultures with a high (0.5 mM) or low (0.2 mM) loading of carbenicillin. All the starter cultures grew well apart from *E. coli* C43 strain, which had no growth.

Starter cultures were used to inoculate flasks of 2TY media and subjected to different growth conditions. A sample of each culture was taken and the cells lysed allowing analysis of the soluble and insoluble fractions to take place (Figure 43). The over expressed DH protein was expected to have a size of approximately 38 kDa. Analysis of the soluble and insoluble fractions by SDS PAGE gel show a band at approximately 38 KDa in both the soluble and insoluble fractions.

Of all the cells screened, the BL21(DE3) codon plus were the most suitable to take forward for further optimisation. The cells grew well, gave good production of the protein in all conditons and tolerated high loadings of antibiotic. As the variations of growth conditions showed no

major difference in the levels of production, the standard growth conditions were taken forward. These conditions were also the easiest to reproduce accurately in future trials. As the cells were able to tolerate the higher antibiotic loading subsequent trials were carried out using an antibiotic loading of 200 µl of 100 mg/ml carbenicillin stock per 200 mL flasks.

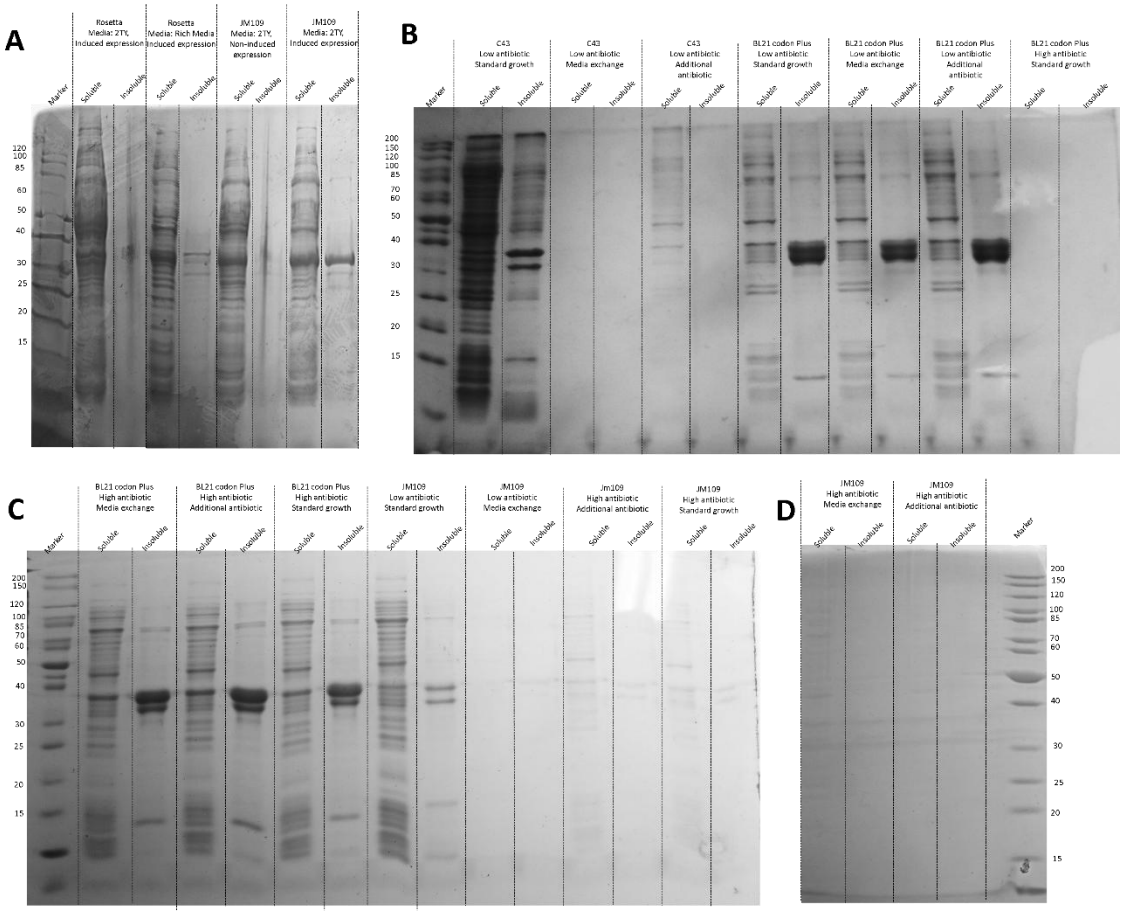


Figure 43 – **A.** SDS PAGE gel showing the levels of soluble and insoluble protein expressed by Rosetta and JM109 *E. coli* strains; **B-D** SDS page gel showing the levels of soluble and insoluble protein expressed by different strains of *E. coli* in 2TY medium under different and growth conditions

3.3.2 Optimisation of expression growth conditions

In addition to poor expression, insoluble protein is a commonly encountered problem during protein over expression in *E. coli*, with more than 30 % of the proteins expressed appearing insoluble.[127] The main purpose of protein expression is to accumulate a significant amount of soluble protein within the cell. Unfortunately, it is common for protein to be misfolded during over expression, leading to insoluble aggregates known as inclusion bodies. Non-optimal conditions during expression can also lead to difficulties when purifying, storing and handling the protein, such as precipitation.[133, 134]

Once the BL21(DE3) codon plus strain had been identified as a suitable expression host, it was possible to optimise conditions to increase the amount of soluble protein expressed by the *E. coli*. The solubility of proteins can be improved by changing expression conditions.

Lowering the expression temperature causes the cell processes to slow down, resulting in a reduction in aggregation of protein.[135, 136] The rates of transcription, translation and cell division are also slowed down resulting in a reduction in the rate of protein synthesis, necessitating the need for longer induction times.[137]

A cheap and easy way of improving the solubility of an expressed protein is to optimise the culture media. In some cases the concentrations of certain salts, peptone and yeast can increase the concentration of soluble protein.[127] Although it has been reported that the increase in protein concentration is due to higher biomass of the culture;[138] rather than the media composition having any major effect on the proteins inherent solubility. Luria broth (LB) is the most commonly used media, composed of bactotryptone, yeast extract and sodium chloride.[136] Terrific broth (TB) has been formulated to increase the yield and solubility of the expressed protein.

Culture conditions also play a large role in achieving high protein expression;[127] the formation of inclusion bodies can be limited or avoided by reducing the rate at which a protein is synthesised. There are a number factors which are tuneable; the cell density before induction can play a major role in achieving good protein solubility. Protein expression is generally induced during the mid-log phase of the cell growth although it can be done at a later stage.

Another way of slowing the synthesis of protein is to tune the level of induction by altering the concentration of inducer used.[125] Although if the concentration of inducer is too low, ineffective induction of expression will result in low yields of protein. There are no hard or fast rules to finding the optimum post induction temperature or duration, so trial and error must be used.

To investigate the effect of media composition on expression levels, BL21 (DE3) codon plus was transformed and grown in a range of different media under induced and non-induced

conditions (Figure 44). Five different media were chosen as potentially suitable candidates; 2TY, 2TY enhanced (2TYE), terrific broth and terrific broth enhanced and Auto induction LB5052. Enhanced media contained additional metals and salts. Induction when needed was induced with 0.4 mM IPTG. Auto induction was not trialled under induced and non-induced conditions as the media contains glucose and lactose. The presence of glucose inhibits expression, whereas the lactose induces expression. The *E. coli* preferentially consume the glucose first, then the lactose. This leads to a gentle gradual induction of expression.

The induction temperature in the standard expression protocol, (Section 7.3) was 19 °C to slow down protein production and improve solubility. The temperature after induction was also decreased from 19 °C to 16 °C in order further decrease the rate of protein production and reduce the chance of misfolded insoluble protein forming.

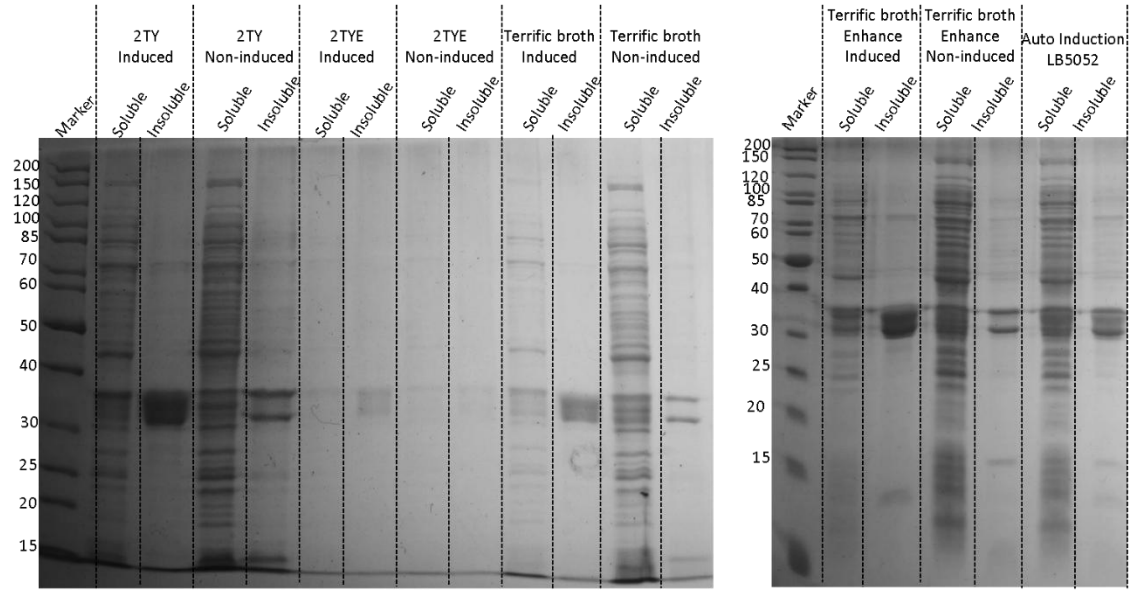


Figure 44 - SDS page gel showing the levels of soluble and insoluble protein expressed by BL21(DE3) codon plus in different media under induced and non-induced conditions,

Protein was produced in 2TY under induced and non-induced conditions, although it was predominantly insoluble in both conditions. A lower amount was expressed by 2TY non-induced as expected. There was no expression in 2TYE and terrific broth enhanced media under induced and non –induced expression conditions. Terrific broth enhanced media gave a large amount of insoluble protein under induced conditions. Under non-induced conditions, terrific broth enhanced had a reasonable amount of soluble expression. Auto induction media was by far the most suitable for over expression of the DH protein; giving a reasonable amount of soluble protein.

The Auto induction medium conditions were taken forward and trialled on a larger scale to assess their suitability for large scale production. A total production culture of 2 litres was grown, all cultures grew well, and the cells were harvested and purified giving a total of 5 mg of

protein. While 2.5 mg of protein per litre is lower than desired, it can be scaled to provide the required amount of protein for assay.

3.3.3 Degradation of protein

During the scale up of Auto induction conditions, four protein bands of substantial size were observed (Figure 45). Two corresponding to the expected weight of the DH protein, approximately 38 kDa, and two approximately 25 kDa in size. Typically, with over expressed proteins, only one band corresponding to the appropriate molecular weight is observed on the SDS page gel in addition to the weaker *E. coli* protein bands. MALDI protein mass spectrometry was carried out on the protein bands in order to identify the origins of the bands (Figure 46, Figure 47, Figure 48 and Figure 49).

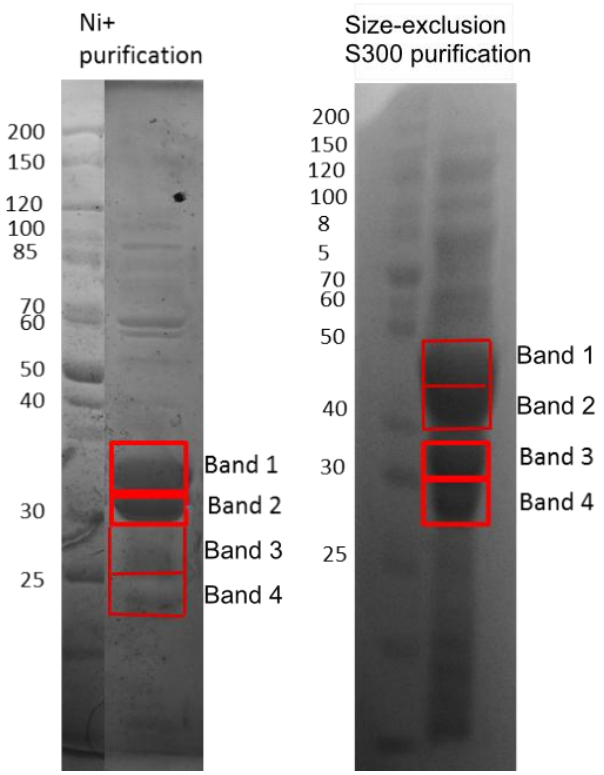


Figure 45 - SDS page gel showing protein bands expression by BL21(DE3) codon plus cells after Ni purification and size exclusion purification. Bands subjected to MALDI analysis, shown in red boxes.

DH band 1 MALDI Result: DH-7615 pOPIN-F Amino Acid Sequence

MAHHHHHHSSGLEVLFQGPNKAHRQRVHPPHDLLGSLIVGR**DLREPTWR**HFIR**VQDIPWIR**DH
VVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVAGYR**LAEVDILR**AMLIPDTSEGLEAHISL
RPCSTKLLLTNEWYDFCVSSVGDDDK**FVDHCR**GR**ITIEFDTSGSADTPR**TSLRERSRSTGLMR**S
VDPSNLYSFLRAQGIYHGPIFQNLK**TISSRKDHSESSFVVANTASVMPNGFQSPHVIHPTTLDSIF
QGAYTALPGAGLDQNTAMIPRSIQELYLSSALTSDVGQCLVSDTSLIRYDGQSFTVNVDVSSK**AD
SEHTPVLEI**KGLRNQSVGQMA

Figure 46 - DH band 1 MALDI Result: DH-7615 pOPIN-F Amino Acid Sequence, found peptide fragments shown in red.

DH band 2 MALDI Result: DH-7615 pOPIN-F Amino Acid Sequence

MAHHHHHHSSGLEVLFQGPNKAHRQRVHPPHDLLGSLIVGR**DLREPTWR**HFIR**VQDIPWIR**DH
VVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVAGYR**LAEVDILR**AMLIPDTSEGLEAHISL
RPCSTKLLLTNEWYDFCVSSVGDDDKFVDHCRGR**ITIEFDTSGSADTPR**TSLRER**SRSTGLMRS
VDPSNLYSFLRAQGIYHGPIFQNLK**TISSRKDHSESSFVVANTASVMPNGFQSPHVIHPTTLDSIF
QGAYTALPGAGLDQNTAMIPRSIQELYLSSALTSDVGQCLVSDTSLIRYDGQSFTVNVDVSSKAD
SEHTPVLEIK**GLRNQSVGQMA**

Figure 47 - DH band 2 MALDI Result: DH-7615 pOPIN-F Amino Acid Sequence, found peptide fragments shown in red.

DH band 3 MALDI Result: DH-7615 pOPIN-F Amino Acid Sequence

MAHHHHHHSSGLEVLFQGPNKAHR**QRVHPPHDLLGSLIVGRDLREPTWR**HFIRVQDIPWIRD
HVVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVAGYR**LAEVDILRAMLIPDTSEGLEAH
ISLRPCSTK**LLLTNEWYDFCVSSVGDDDK**FVDHCRGRITIEFDTSGSADTPR**TSLRER**SRSTGL
MRSVDPSNLYSFLRAQGIYHGPIFQNLK**TISSRKDHSESSFVVANTASVMPNGFQSPHVIHPTTL
DSIFQGAYTALPGAGLDQNTAMIPRSIQELYLSSALTSDVGQCLVSDTSLIRYDGQSFTVNVDVSS
KADSEHTPVLEIKGLRNQSVGQMA

Figure 48 - DH band 3 MALDI Result: DH-7615 pOPIN-F Amino Acid Sequence, found peptide fragments shown in red.

DH band 4 MALDI Result: catabolite activator [Escherichia coli] 23.77 kDa

MVLGKPQTDPTLEWFLSHCHIHKYPSKSTLIHQGEKAETLYYIVKGSVAVLIKDEEGKEMILSYL
NQGDFIGELGLFEEGQERSAWVRAKTACEVAEISYKKFRQLIQVNPDILMRLSAQMARRLQVTS
EK**VGNLAFLDVTGRIAQTLLNLAK**QPDAMTHPDGMQIKITR**QEIGQIVGCSR**ETVGHILK**MLE
DQNLISAHGKTIVVYGTR**

Figure 49 - DH band 4 MALDI Result: catabolite activator [Escherichia coli] 23.77 kDa, found peptide fragments shown in red.

Analysis of MALDI spectra for each band allowed the identification of each protein. Fragments of peptide corresponding to the DH domain were found in three of the four bands. The above figures (Figure 46, Figure 47, Figure 48, Figure 49) show the amino acid sequence and the peptide fragments found for each protein band identified.

As three of the four bands of peptide fragments correspond to the DH protein, the DH protein must have been undergoing degradation at some point during the growth and purification process. The DH band 1(Figure 46) appears to be the full length protein; whereas DH band 2 (Figure 47) appears to be truncated at the N-terminus as the end sequence fragment is present. It is unclear whether the DH band 3 is truncated at the N or C terminus due to the fragmentation pattern (Figure 48). DH band 4 is an unrelated *E. coli* band (Figure 49).

The degradation of protein was most likely to be occurring during purification. Although the BL21 (DE3) codon plus cells are deficient in *lon* and *ompT* proteases, when the cells are lysed the protein can come into contact with other proteases present within the cell. If the degradation was being caused by additional proteases present within the cell, the use of a protease inhibitor cocktail during purification would prevent this degradation from occurring.

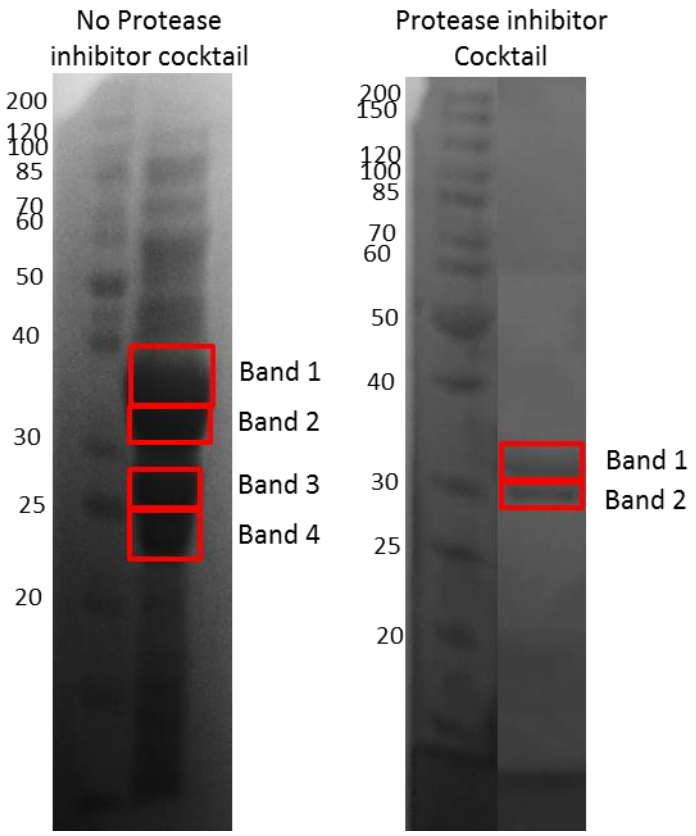


Figure 50 – SDS page gel of protein expressed and isolated from BL21(DE3) codon plus without the presence of a protease inhibitor cocktail and SDS page gel of protein expressed and isolated from BL21(DE3) with the presence of a protease inhibitor cocktail.

Although the use of a protease inhibitor cocktail decreased the level of degradation in the expressed protein (Figure 51), it did not fully eliminate it and two bands of DH protein were still observed. Alternative cell strains (BL21 (DE3), BL21 (DE3) codon plus, KRX) were investigated (Figure 51). BL21 (DE3) and BL21 (DE3) codon plus are both deficient in *lon* and *ompT* proteases whereas KRX single step cells are deficient in *ompP* and *ompT* proteases.[139] Over expression of protein in KRX is induced by the addition of rhamnose as the strain processes a chromosomal copy T7 RNA polymerase gene which is driven by a rhamnose promoter.

Protein was over expressed by all three strains, BL21 (DE3) and BL21 (DE3) codon plus strains had higher levels of expression but the degradation of protein persisted in these strains. Although the level of expression by the KRX cells was very low, there appeared to only be one band of DH protein present in the soluble and insoluble fraction.

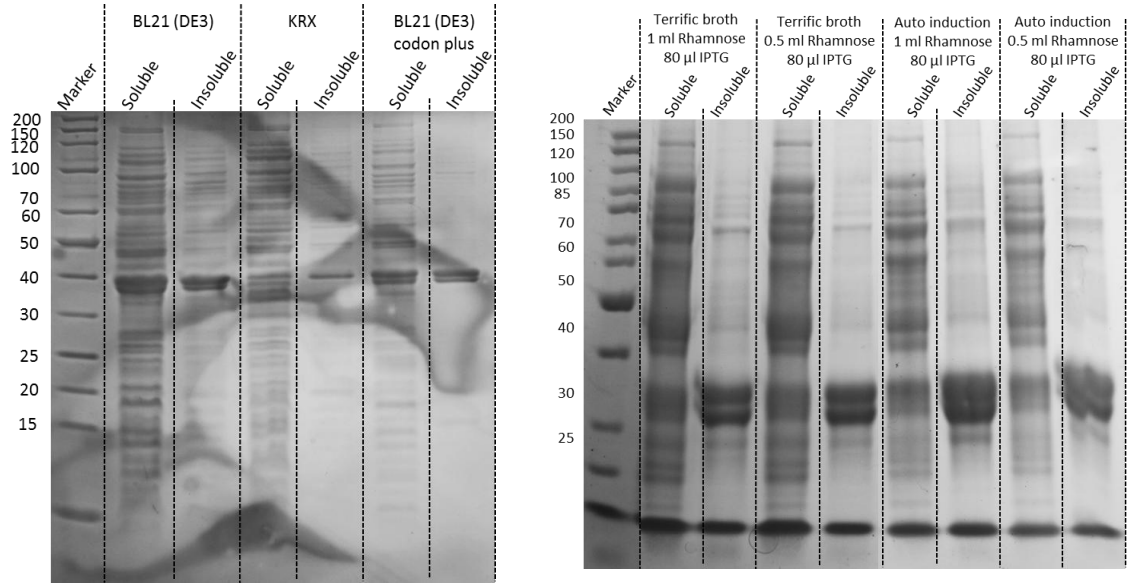


Figure 51 – **A,** SDS page gel showing the levels of soluble and insoluble protein expressed by protease deficient cell strains: BL21(DE3), BL21(DE3 codon plus and KRX ; **B**, SDS page gel showing the levels of soluble and insoluble protein expressed by KRX during optimisation of growth conditions.

As degradation of the protein occurred in the BL21 (DE3) and BL21 (DE3) codon plus strains, further optimisation of the KRX growth conditions was undertaken investigating the effect of media and induction conditions (Figure 51, B). IPTG was also added to all growth cultures as recommended by the technical literature supplied by Promega. The DH expression vector uses a *lac* operator which is adjacent to the T7 promoter, therefore IPTG may be required to achieve maximal expression. KRX cell expressed well in all media conditions, however the majority was insoluble, and degradation of the protein had occurred in both the soluble and insoluble phase. Subsequent trials utilising lower concentration of rhamnose were carried out to slow down the rate of protein production and therefore aid solubility. However, it was found upon induction that the protein expressed was toxic to the cells.

A time course was carried out using the BL21 (DE3) strain to give more information as to when the protein degradation was occurring. Samples of the culture were taken at regular time points and the crude, soluble and insoluble fractions analysed by SDS page gel.

After 4 hours, over expression of the DH protein starts to occur in the soluble fractions although there are greater amounts of protein in the insoluble phase, which had undergone degradation (Figure 53). Degradation was observed in the soluble fraction after nine hours of incubation.

In a final attempt to overcome the degradation of protein during expression, a trial using the *E. coli* BL21 Gold (DE3) strain was carried out. BL21 (DE3) was also used as a comparison. This was grown to stationary phase overnight at 30 °C with shaking. As the BL21 Gold (DE3) strain had not been grown before, different media (auto inductions, LB) and growth conditions (induced and non-induced) were used (Figure 52). It was found that protein was expressed by BL21 Gold (DE3) cells under all conditions. It was observed that protein degradation was less apparent in the BL21 Gold (DE3) strain when grown in LB media. A large scale production culture of 4.8 litres was carried out using BL21 Gold (DE3). The production culture grew well, and the cells were harvested and purified. A total of 44 mg of pure, soluble DH protein was isolated, while 9 mg/L is not an ideal quantity, the production is scalable to allow adequate amounts of protein to be produced for the required analysis.

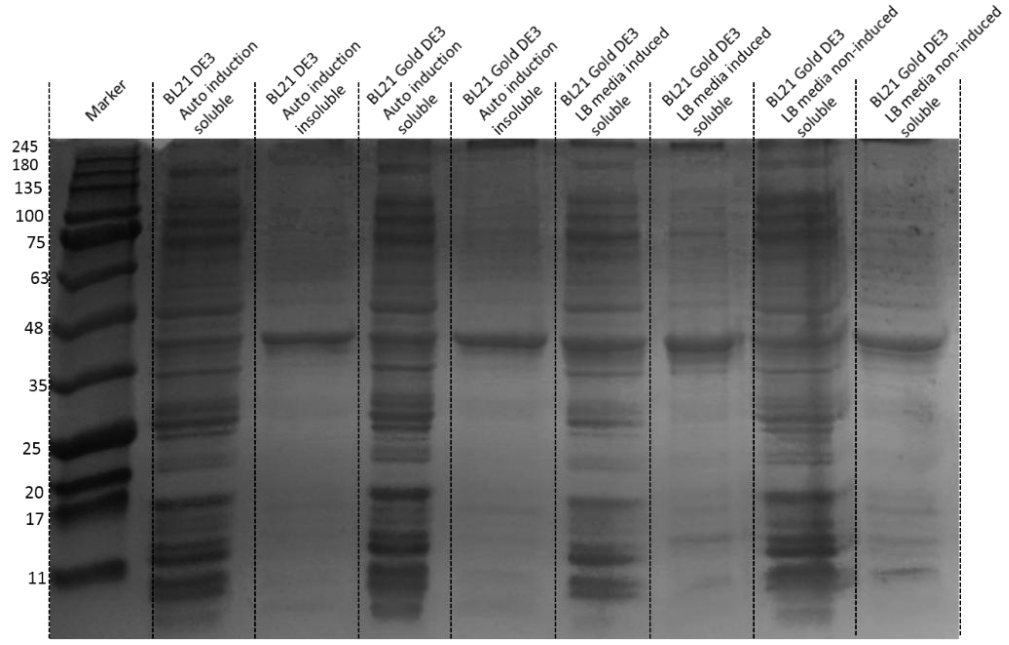


Figure 52 - SDS page gel of BL21 Gold (DE3) trial.

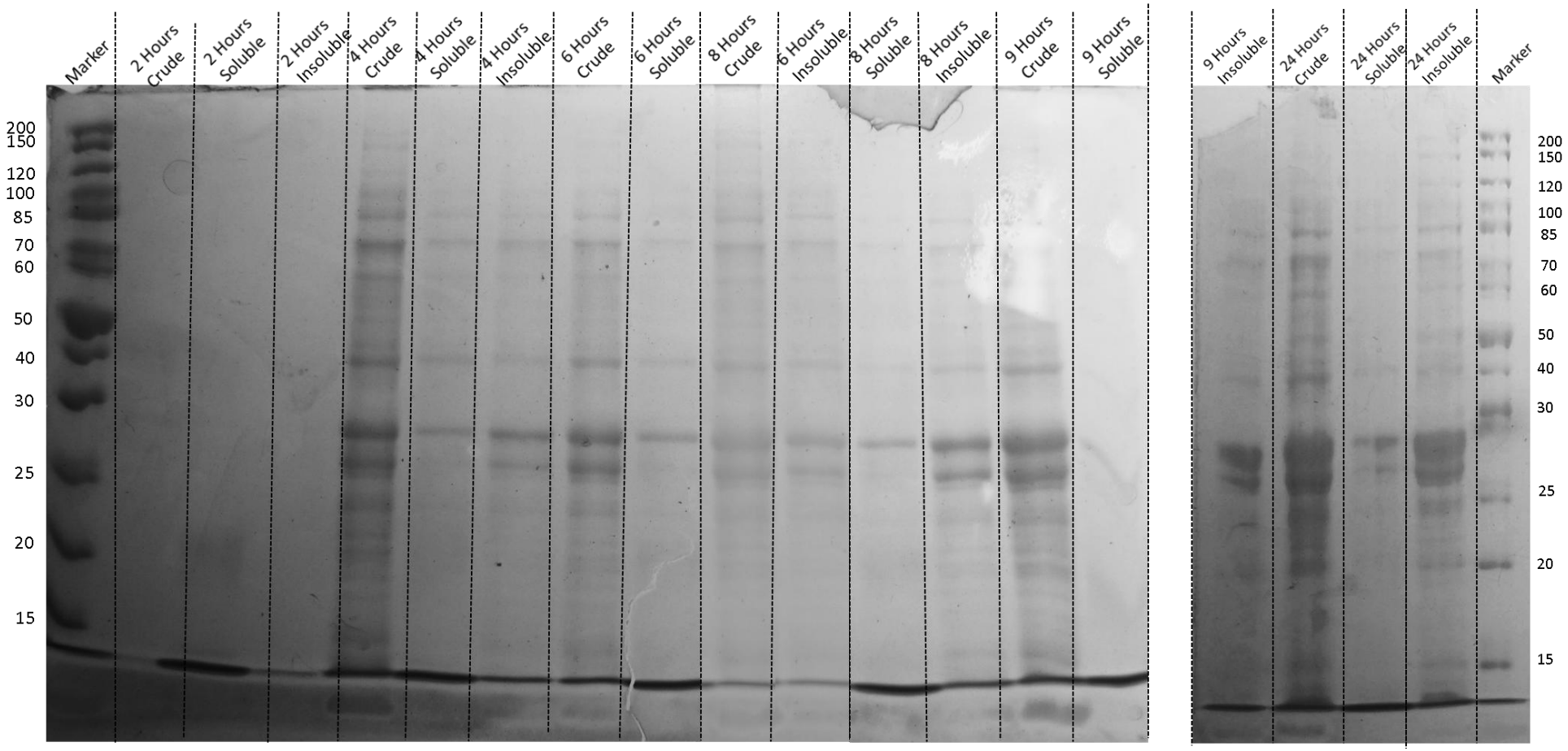


Figure 53 - SDS page gel showing the levels of soluble and insoluble protein expressed by BL21(DE3) over the course of 24 hours.

3.3.4 Optimisation of buffers and purification protocol

The plasmid #7615 designed by David Ivision included a vector coding for N-terminal hexahistidine tags (His-Tag).[95] The presence of the His-Tag allows the over expressed protein to be easily purified from the bulk of the *E. coli* proteins.[140] This is achieved due to the high affinity of the His-Tags for Ni^{2+}. The protein is purified by passing the lysed cells through a nickel-sepharose high performance column, the His-Tag of the over expressed protein of interest interacts with the metal ions in the column matrix causing it to bind. The bulk bacterial proteins do not bind and are not retained on the column. Once the unwanted protein has been washed though the column using a binding buffer, the protein of interest can be eluted by a solution of imidazole in an elution buffer. Further purification can be carried out using a FPLC S300 size exclusion column to separate the proteins by size.

A number of issues were encountered during the optimisation of protein purification. The degradation of protein during purification required the addition of a protease inhibitor cocktail to the loading buffer to limit the extent of the degradation (Section 3.3.3).

Achieving the desired level of purity from the Ni^{2+} column required some alterations in protocol from that used by David Ivision.[95] Initially a ready to use 5 mL HisTrap Fast Flow column from GE was being used. Although the column was reported to have a high binding capacity, large amounts of soluble protein were lost from the Ni^{2+} column. A larger Ni^{2+} sepharose fast flow column of 25 mL was prepared using Ni^{2+} sepharose 6 fast flow BioProcess medium. This increase in size allowed greater volumes of lysed cell to be passed though the column without losing the desired over expressed protein due to overloading.

While increasing the loading capacity of the Ni^{2+} column reduced the amount of over expressed protein from being lost, it also increased the amount of unwanted bacterial protein retained on the column. Increasing the concentration of imidazole present in the binding buffer from 5 mM to 20 mM dramatically increased the level of purity obtained. A high enough level of purity was obtained to make it unnecessary for any further purification via S300 size exclusion.

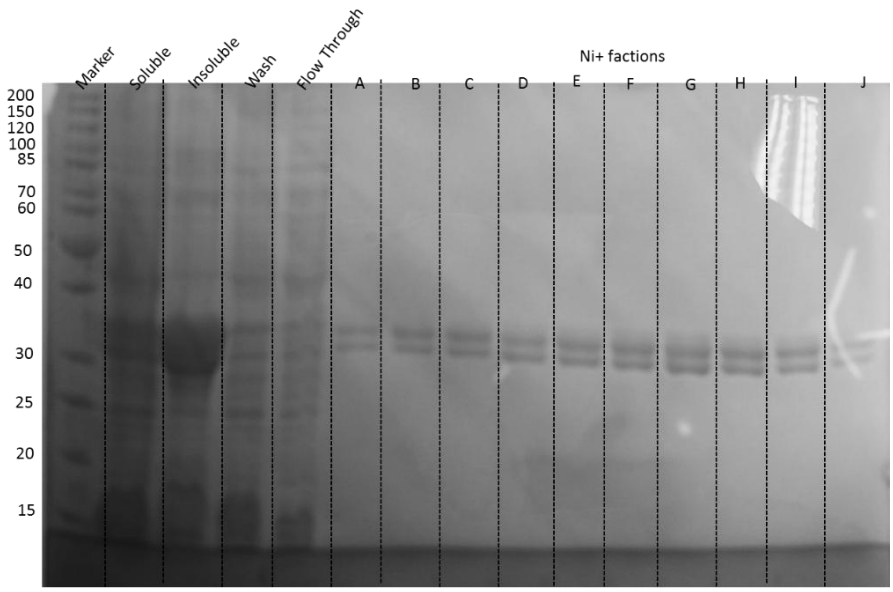


Figure 54 – Purification factions using 25 mL Ni^{2+} sepharose fast flow column

Once pure protein has been produced in a soluble form, concentrating it to the desired level without causing precipitation or aggregation can be highly problematic and a common stumbling block.[134] It is estimated that 33 – 55 % of all expressed non-membrane proteins are not soluble and up to 50 % of the remaining soluble proteins aggregate or precipitate during concentration. During purification of the DH protein, the problem of precipitation was encountered. It was found that after elution from the Ni column, the protein tended to precipitate immediately or soon after.

It was noticed that the protein had a low tolerance for high concentration of imidazole. This issue was overcome by eluting the His-Tag protein slowly using a gradient change in the concentration of imidazole, thus eluting the protein at lower imidazole concentrations. After the protein had been eluted from the Ni column, it was concentrated and passed through a desalting column to remove the imidazole and reduce the risk of precipitation caused by high salt levels.

Alongside the change in elution method, the buffer conditions were screened to establish which afforded the best solubility.[134] Previously the buffers were composed as follows; Binding buffer contained 20 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM imidazole, 10 % glycerol; Elution buffer contained 20 mM Tris-HCl pH 8, 150 mM NaCl, 400 mM imidazole, 10 % glycerol; Desalting buffer contained 20 mM Tris-HCl pH 8, 150 mM NaCl, 10 % glycerol.

It is widely accepted that addition of additives, such as ionic compounds, salts detergents etc. can increase the solubility and stability of protein.[141] A simple screening of buffers and additives was carried out using a hanging drop experiment using a 16 well plate. The buffer and additives detailed in the JBS solubility kit available from Jana bioscience were used as guide for selecting possible buffers and additives to screen.[141]

pH buffers ranging from pH 4-10, salt concentrations, buffer type and some additives were screened (Table 1). The DH protein has a predicted isoelectric point of 6.93, therefore buffers with a pH close to pH 7 should be avoided. The wells were inspected regularly over a 24-hour period.

Buffer	Additive
Sodium acetate pH 4.7	40 % Glycerol
Sodium phosphate pH 5	10 mM DTT
Potassium phosphate pH 6	1 M NaCl
Tris HCl pH 8.5	100 mM L-Arg & L-Glu
CHES pH 9	
CHES pH 10	

Table 1 - Buffers and additives used in initial screen.

After 24 hours high level of precipitation was observed across all buffers in wells containing 40% glycerol and 1M NaCl, making these two additives the least suitable.

At pH 8.5 and pH 10 DTT was not suitable with precipitation occurring very quickly, within the first hour. However, at pH 4 and 5, protein precipitation was much slower with visible quantities of protein forming after 4 hours. 100 mM L-Arginine and L-Glutamic acid was found to be the most suitable additive, improving the proteins solubility and stability at pH 4, 5 and 8.5, with no precipitate forming after 24 hours. It was decided that the most suitable buffer composition for protein storage were as follows: binding buffer:10% glycerol, 50 mM Tris HCl pH 8, 150 mM NaCl and 100 mM L-Arginine and L-Glutamic acid per L.

3.4 Crystallisation

David Ivision attempted to crystallize the DH protein using four commercial crystallisation screens: Structure screen 1 + 2HT-96, JCSG+ HT96, PACT premier HT-96 and Morpheus HT-96 (all from Molecular Dimensions).[95] Sitting drop crystallisation trays were set up using an Art Robbins Pheonix crystallisation robot and incubated at 20 °C for several days. Inspection of the trays showed heavy precipitate but no defined crystals.

The four commercial crystallisation screens were repeated using DH protein stored in the optimised buffer. Sitting drop crystallisation trays containing 50 µl well solution and 0.1 µl well solution + 0.1 µl protein solution were set up using an Art Robbins Pheonix crystallisation robot and incubated at 20 °C. After 7 days the trays were inspected and showed precipitation in most drops. No further attempts to crystallise the DH domain were undertaken.

3.5 Conclusion

After optimisation of the expression conditions, it was possible to isolate approximately 10 mg/L of soluble active SQTKS DH protein. The optimisation of the purification and storage

buffers allowed the purified DH protein to be stored at 4 °C for up to 2 months without precipitation.

The optimised protocol was as follows: *E. coli* BL21 Gold (DE3) was used as a host for protein expression. Following transformation with the appropriate expression plasmid using LB-Glu/Car agar plates, a single colony was used to inoculate a 50 mL sterile LB starter culture medium containing 0.2 mM carbenicillin. This was grown to stationary phase overnight at 30 °C with shaking. This starter culture was used to inoculate LB medium supplemented with carbenicillin at 1:100 dilution in 500 mL flasks each containing 200 mL of LB medium. The flasks were incubated at 37 °C with shaking till an OD_{600} of 0.4 was reached. The temperature of the culture was then adjusted to the expression temperature of 16 °C. Induction when needed was induced with 0.4 mM IPTG. The cultures were incubated with shaking for a further 6 hours and cells were harvested by centrifugation. The cell pellets suspended in binding buffer containing protease inhibitor cocktail and were refrigerated overnight before being purified the following day. The full protocol for purification is detailed in chapter 7, Section 7.4.

Chapter 4 - Investigating the Selectivity and Kinetics of the Isolated SQTKS DH Domain

4.1 Introduction

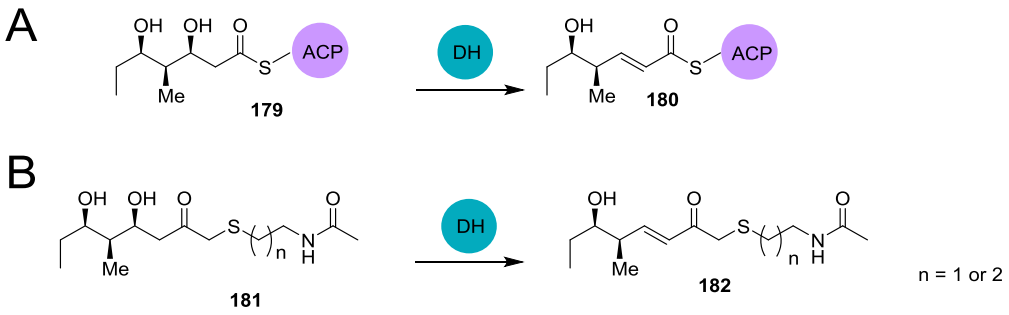
While it is possible to carry out assays using acyl-ACP species as substrates for PKS enzymes, there are many problems associated with their use.[104] One such difficulty is producing a sufficient quantity of ACP to carry out a kinetic assay.[104] Also assays involving ACP bound substrate mimics typically require a TE domain or complex time consuming procedure to release the enzymatic product.[104] As there is no release activity observed for the SQTKS system stoichiometric amounts of ACP would be required. For example, carrying out kinetic assays utilizing the SQTKS ACP would require 150 - 200 mg of protein per substrate. This amount is calculated assuming the K_M is 1 mM and the assay volume is 50 µL. The SQTKS ACP is 10 KDa, and assuming the K_M is 1 mM, results in the need to test concentrations of substrate up to 10 mM, and would result in a maximum of 5 mg of protein per assay. Typically kinetic assays are repeated in triplicate to ensure accuracy and repeatability. Previous attempts to produce the SQTKS ACP by David Ivision resulted in no soluble protein.[95]

While the substrate specificity of many PKS DH domains has been interrogated using ACP bound substrate mimics, simple diffusible substrates have also been used. Studies into Rif10DH[13] and Ery4DH[12] utilised ACP bound mimics as well as simple diffusible substrates. The investigation into Pikromycin DH (Pik2DH) carried by Wu *et al.* utilised only diffusible substrates.[65] *N*-Acetylcysteamine (SNAC) or pantetheine based substrate mimics have also been used in multiple studies of other PKS enzymes.[13, 65, 67] These simple mimics are able to diffuse into the active site of the domain.[67, 104, 142] While using diffusible substrate mimics has its advantages, there are also some disadvantages; for example the protein-protein interaction between the ACP and the catalytic domain might be vital to achieving substrate fidelity.[65]

Work by Keating-Clay and co-workers investigating the stereoselectivity of the rifamycin DH, Rif10DH, highlighted the impact protein-protein interactions have on substrate fidelity (fully discussed in Chapter 1, Section 1.5.7).[13] The stereoselectivity of Rif10DH altered depending on whether the substrate was diffusible or attached to the native or a non-native ACP. When diffusible SNAC or pantetheine (Pant) based substrates were incubated with the Rif10DH the DH domain exhibited high selectivity towards substrates with 2*R*,3*R* stereochemistry **73** and **74**. When the substrate was bound to the native ACP, Rif10ACP selectivity favoured the 2*S*,3*S*-2-methyl-3-hydroxypentanoyl stereoisomer **83**. When bound to the non-native ACP, Ery6ACP, the stereoselectivity again reversed and only the 2*R*,3*R*-2-methyl-3-hydroxypenanoyl-Ery6ACP **88** substrate was accepted.

While diffusible substrate mimics have been successfully used to interrogate the substrate specificity of excised DH domains,[13, 65, 67] very little work has been carried out using these mimics to investigate the steady state kinetics of isolated DH domains.[67] Typically isolated DH domains have low activity and therefore require long assay incubation times in order to generate sufficient product for detection.[13, 67, 104]

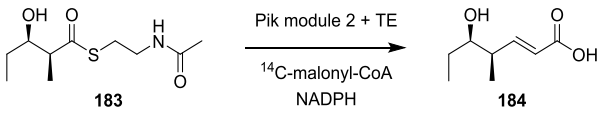
Li *et al.* utilised liquid chromatography-tandem mass spectrometry (LC-MS/MS) to carry out steady state kinetics on the pikromycin DH domain, Pik2DH.[67] Their previous work carried out on the Pik2KR domain showed that LC-MS/MS was highly selective and sensitive to DH substrates and products and therefore was deemed suitable for the analysis of DH substrate turnover.[67, 104] A multi-domain construct was built, (Pik2KR-DH) in order to achieve a more native environment for the assay. After overexpression and purification of the protein a range of assays were carried out in order to probe the selectivity of the active site.



Scheme 55 – **A**, Natural PikDH2 substrate **179** and product **180**; **B**, rational design of PikDH2 substrate **181** mimics to inhibit lactonisation and product **182**.

The simple diffusible SNAC based substrate **181** was used for the assay (Scheme 55, B). The substrate mimic was based on the natural product triketide **179** with two differences. First, the native ACP-phosphopantetheine arm was replaced by SNAC; second, either a one or two carbon methylene spacer was introduced between the carbonyl and the sulfur group to inhibit lactonisation.

Work carried out by Wu *et al.* examining the kinetic characterisation of dehydratase containing pikromycin module 2 (Pik2) utilised TLC phosphor-imaging to quantitatively measure substrate turnover.[65] The aim of the study was to investigate the kinetic selectivity of Pik2DH when part of a complete module rather than as an isolated domain. The diketide substrate **183** was incubated in the presence of malonyl-CoA, and the module was able to carry out chain extension (KS), reduction (KR) and dehydration (DH) reactions to form **184** (Scheme 56).

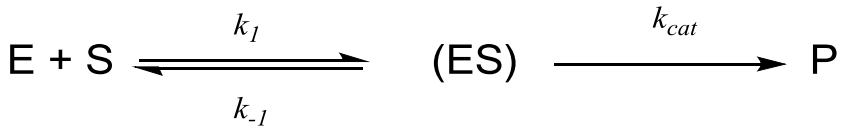


Scheme 56 – Enzymatic conversion of diketide **183** to triketide **184** by Pik module 2 + TE

The kinetic characterisation assay required diketide substrate **183** to be incubated at 30 °C for 40 minutes in the presence of ^{14}C-malonyl-CoA, NADPH and Pik2 + a TE domain. The reaction was quenched by adding 1 M HCl and subsequently extracted using ethyl acetate (3 x 180 µl). The extract was dried and redissolved in ethyl acetate (30 µl), spotted on to TLC plates and developed using dichloromethane, 5% methanol and 0.1% acetic acid. The samples were analysed by phosphor-imaging, using a known concentration of ^{14}C-malonyl-CoA as a standard. The use of phosphor-imaging is ideal for monitoring this type of reaction where a whole module is being investigated, as chain extension allows the installation of ^{14}C. Phosphor-imaging is not suitable for studying a DH domain in isolation, as the dehydration reaction only results in the loss of water, so radio-labels cannot be incorporated.

4.2 Michaelis-Menten kinetics

The model which describes the rate of enzyme catalysed reactions was developed by Michaelis and Menten in 1913.[143] Their work on invertase showed that the rate of an enzyme catalysed reaction is proportional to the concentration of the enzyme-substrate complex. The model developed by Michaelis *et al.* makes two assumptions: first the enzyme (E) and substrate (S) make a reversible complex (ES); and, second, that the reversible complex (ES) irreversibly yields the product (P, Scheme 57)



Scheme 57 – The progress of a Michaelis-Menton enzyme catalysed reaction.

The rate at which a substrate associates and binds to an active site is described by k_1. The dissociation rate is described by k_{-1}. Together these two rate constants describe how well a substrate binds to the active site. Substrates with a larger k_1/k_{-1} value will bind more strongly to the active site.

k_{cat} is usually the rate limiting step, by which the enzyme transforms the substrate into the product (P). The efficiency of an enzymatic reaction is defined by the two parameters k_1/k_{-1} and k_{cat}. The Michaelis-Menten equation (Eq. 1) is derived from this simple model by extrapolating the equations. In order to do this the steady state approximation is applied, this approximation assumes that the concentration of the ES complex is constant.

$$v_0 = \frac{V_{\max}[S]}{K_M + [S]} \qquad \text{Eq. 1}$$

Where v_0 is the initial rate of reaction, obtained when there has been no appreciable decrease in substrate or formation of product. V_{max} is the maximum rate of reaction the enzyme can achieve. This is where the concentration of substrate is so great that all the enzyme is in the ES state. The final term K_M is the Michealis Constant, which represents the substrate concentration when the rate of reaction is half the rate of V_{max} (Figure 55).

When the enzyme concentration is kept constant and the concentration of the substrate increases, the overall rate v_0 will increase in a near linear fashion as long as the substrate concentration is well below K_M. As more substrate is added, there is more substrate available to bind to the active site of the enzyme and therefore the concentration of the enzyme-substrate complex (ES) increases.

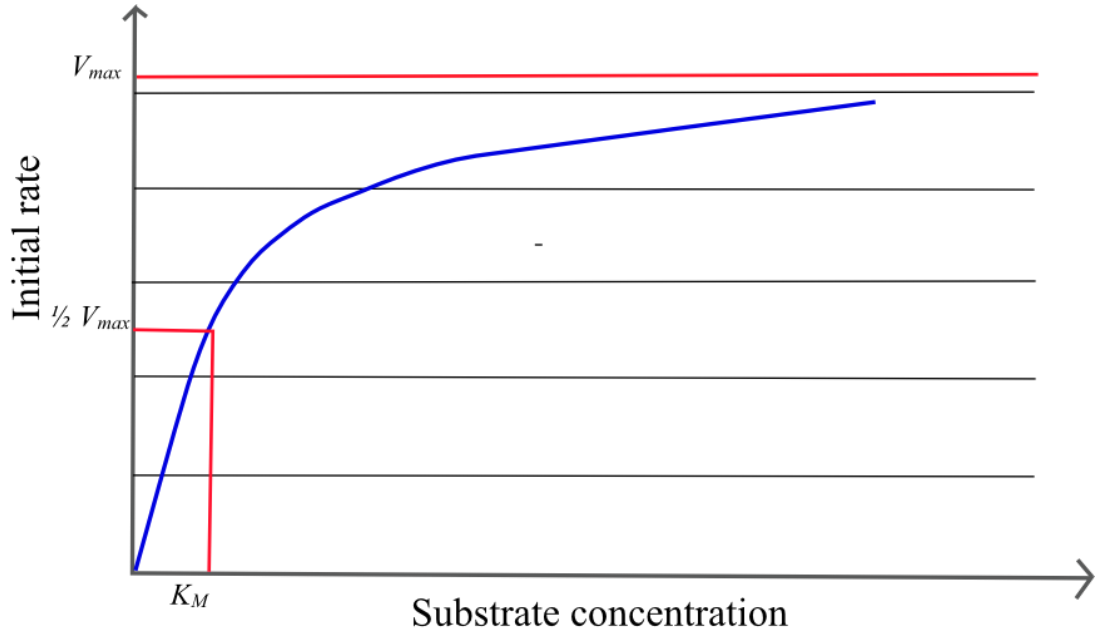


Figure 55 – Plot of initial rate against substrate concentration in a theoretical enzyme reaction demonstrating Michaelis-Menton kinetics.

At very high concentrations of substrate, much greater than K_M, there is less free enzyme in solution due to the substrate binding to the enzyme to form the ES complex and the increase in the rate is no longer linear. As previously stated, when the rate of reaction is at V_{max}, the substrate concentration [S] is so high there is no free enzyme, as all the enzyme is occupied in a ES complex. It is therefore possible to describe V_{max} as follows in Eq. 2. Substrates which bind strongly to the active site will have a higher concentration in solution, therefore they will approach V_{max} at a faster rate to those which bind more weakly.

$$V_{\max} = [E_0]k_{\text{cat}} \quad \text{Eq. 2}$$

It is possible to define the concentration of substrate and ES as follows:

$$[E].[S] = [ES].\frac{k_{-1} + k_{\text{cat}}}{k_1} \quad \text{Eq. 3}$$

The Michealis constant, K_M, is the substrate concentration at which the rate is half of V_{max}. It is therefore possible to define K_M in a way which is directly relatable to k_1 and k_{-1} (Eq. 4).

$$K_{\text{M}} = \frac{k_{-1} + k_{\text{cat}}}{k_1} \quad \text{Eq. 4}$$

K_M describes the affinity a substrate has for binding to an enzyme. A substrate with poor binding affinity would result in a higher K_M value. A substrate with a high binding affinity would have a low K_M value. While K_M describes the binding affinity of a substrate, k_{cat} describes, the second stage of the enzyme reaction, how quickly the substrate is converted to the product. Using these two parameters it is possible to describe how efficiently a particular substrate is processed by a particular enzyme.

$$Substrate\ efficiency = \frac{k_{\text{cat}}}{K_{\text{M}}} \quad \text{Eq. 5}$$

The enzymatic efficiency had been investigated in a number of DH domains. In order to investigate the stereoselecity of the SQTKS DH a set of assays were undertaken to measure k_{cat} and V_{max} for each of the possible four substrates.

4.3 Results

4.3.1 Assay development

4.3.1.1 Previous work on SQTKS DH domain

Initial work by David Ivison, (Section 1.6.2), produced SQTKS DH as a standalone protein.[95] The newly isolated DH domain (132 µM) was incubated with a racemic mixture of stereoisomers 2-methyl-3-hydroxybutanoyl-SNAC **111** (0.5 mM) in tris buffer (pH 7) overnight. The control assay contained only substrate and buffer. The reactions were quenched and analysed by LCMS. The resulting LCMS trace showed an additional peak at 5.2 mins, the mass spectrum for this time corresponded to that of tigloyl-SNAC **112**, *m/z* 224 [M] Na^+ (Figure 56).

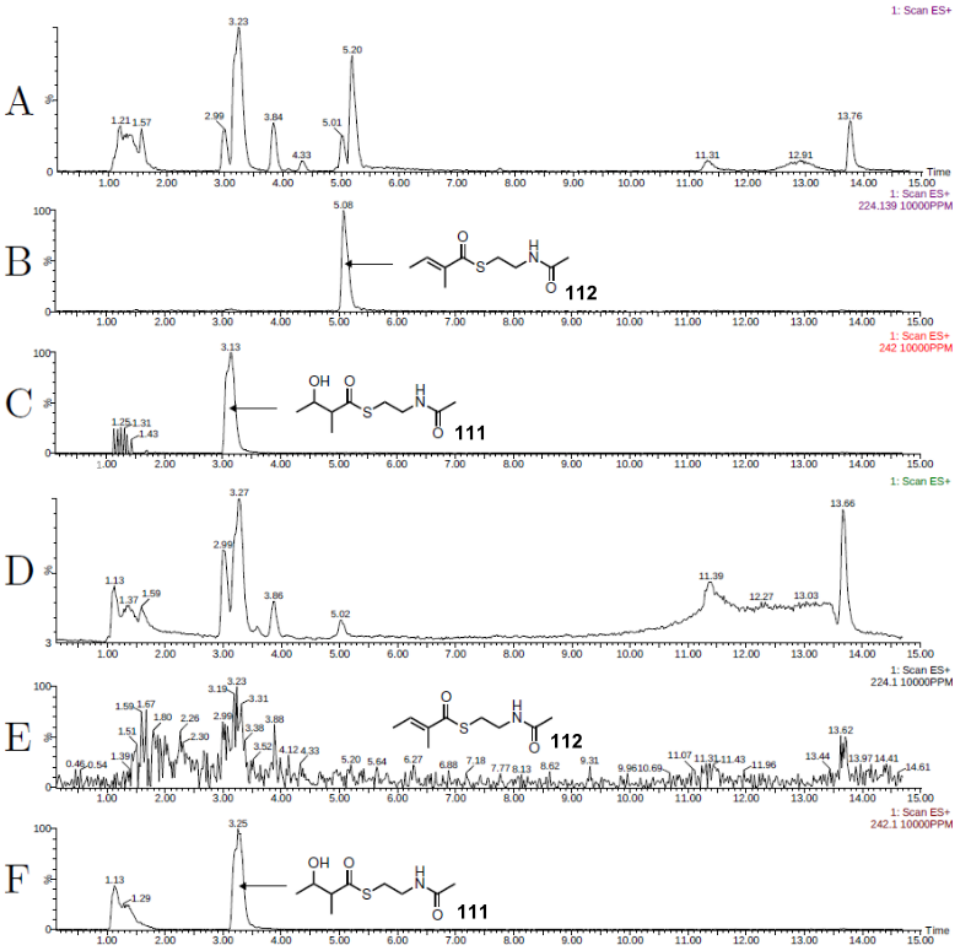


Figure 56 – Assay carried out by David Ivision: [95] **A**, DH assay ESI$^+$ – TIC chromatogram; **B**, DH assay ESI$^+$ EIC (tigloyl-SNAC **112**); **C,** DH assay ESI$^+$ EIC (3-hydroxy-2-methylbutyryl-SNAC **111**); **D,** control ESI + TIC chromatogram; **E,** control ESI + EIC (tigloyl-SNAC **112**); **F,** control ESI + EIC (3-hydroxy-2-methylbutyryl-SNAC **111**). EIC = extracted ion chromatogram

4.3.1.2 Substrates synthesised

Work carried out by David Ivison on the SQTKS DH proved that the product of the dehydration is tigloyl-SNAC **112**.[95] However the stereoselectivity of the DH was still unknown, as racemic 2-methyl-3-hydroxybutanoyl-SNAC **111** was used in the assay.

In order to probe the stereoselectivity of the DH a range of diastereomeric and enantiomeric diketide-SNAC substrate were synthesised (Chapter 2). In total there were 4 natural substrate mimics **114**, **115**, **116** and **117** (Figure 57). Two non-natural substrate mimics were also synthesised, **124** and **125**. The *E*- DH product, **112** was also synthesised as a standard.

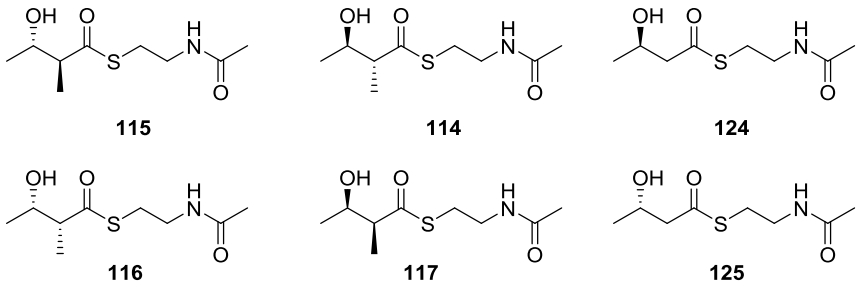


Figure 57 – diketide substrates synthesised in order to probe the DH active site.

4.3.1.3 Substrate retention time on LCMS

In order to monitor the dehydration reaction carried out by the DH domain, standards of the substrate and product mimics were analysed using LCMS. This allowed the suitability of LCMS assay monitoring to be assessed and the specific retention time and fragmentation patterns to be observed. All substrate mimics **114**, **115**, **116** and **117** were found to have similar retention times ranging from 3.9 – 4.1 minutes. The standard of the enzymatic product **112** was found to have a retention time of 5.5 minutes (Figure 58).

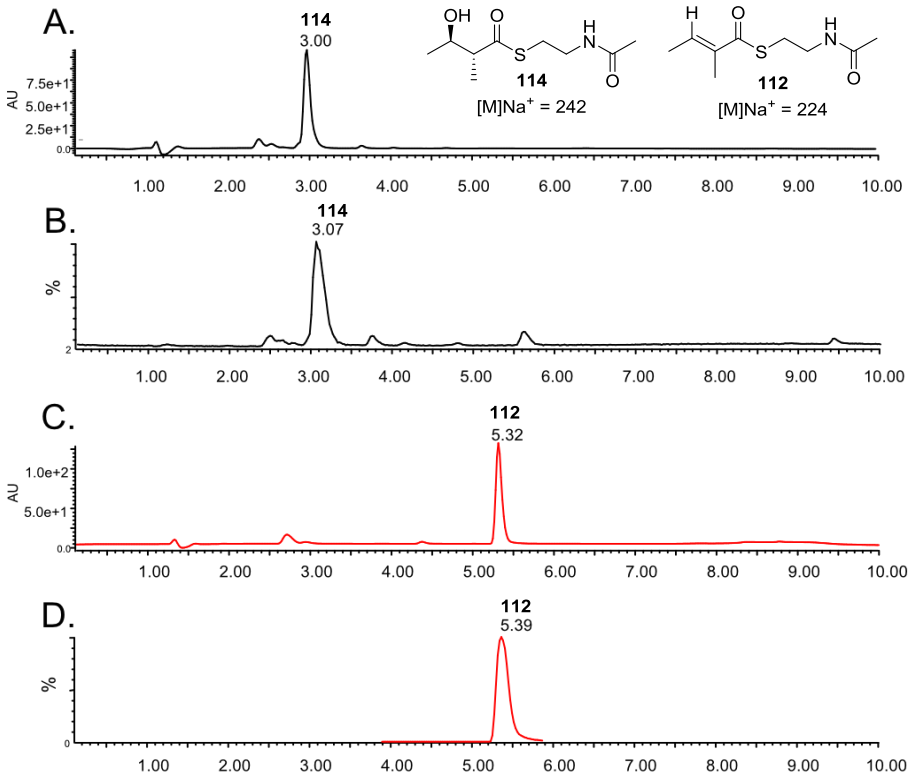


Figure 58 – **A**. UV trace of 2*R*, 3*R*-**114**; **B,** ESI+ Mass trace of 2*R*, 3*R*-**114**; **C**, UV trace of **112**; **D**. ESI+ Mass trace of **112**.

4.3.1.4 Initial assay

After the protein expression process was optimised, (Chapter 3) the assay carried out by David Ivison was repeated. The freshly prepared DH (100 µL, 132 µM) was incubated with substrate 2*R*,3*R*-**114** (1.7 mM, 10 µL) and protein storage buffer pH 8 (1 M Tris-HCl pH 8 (10 µL), 50 % Glycerol (40 µL)) overnight at 30 °C. The reaction was quenched with MeCN:H_2O (60:20 µl) and the precipitated protein was removed *via* centrifugation. The supernatant was then analysed directly by LCMS.

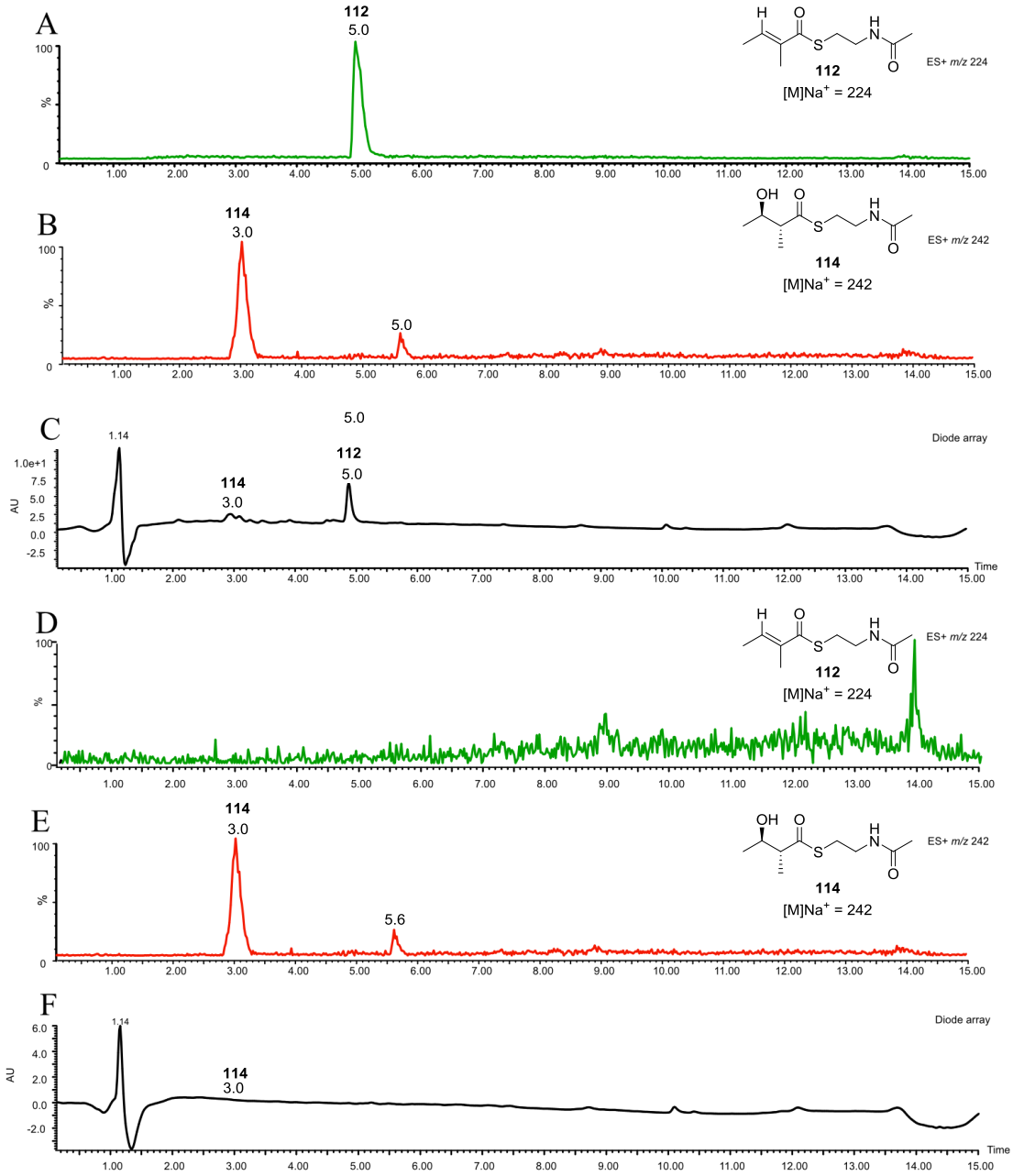


Figure 59 – A, DH assay, single ion EIC $^+$ trace, **112** $[M]Na^+$ = 224; **B,** DH assay, EIC^+ **114** $[M]Na^+$ = 242; **C,** DH assay, UV trace of initial assay, **D,** Control assay, single ion EIC $^+$ trace, **112** $[M]Na^+$ = 224, **E,** Control assay EIC $^+$ trace of **114** $[M]Na^+$ = 242, **F**. Control assay, UV trace of initial assay.

LCMS analysis of the DH assay (Figure 59, C) shows the enzyme is active and the substrate has been dehydrated to form the expected product. In the DH assay (Figure 59,C) there is one peak with a retention time of 5.0 minutes. This peak was confirmed to be the product **112** by interrogation of the mass spectrum (Figure 60, A, *m/*z 224 $[M]Na^+$) and comparison to the standard. Traces of the substrate **114** were shown to be present in the EIC with a mass of *m/z* 242 $[M]Na^+$, retention time 3.0 mins (Figure 59,B).

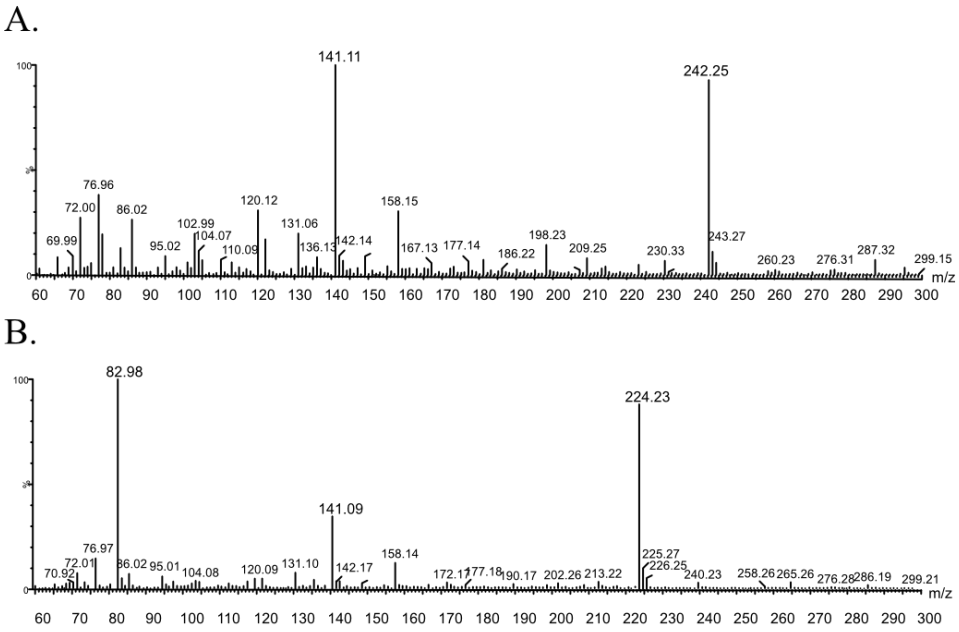


Figure 60 – A, ESI+ of substrate **114** $[M]Na^+$ = 242**; B,** ESI+ of product **112** [M] Na^+ = 224.

Control assays used identical conditions, and this ensured the concentration of all other components stayed the same. Ideally, control assays would have been carried out using inactive or denatured protein. Although it was possible to denature the DH protein with heat, it was not possible to keep the denatured protein in solution. As soon as heat was applied, the protein precipitated. Therefore, in control assays the protein was replaced with protein storage buffer. LCMS analysis of the control assay did not show the formation of new peaks (Figure 59, E). A peak at 3.0 minutes with a mass spectrum of *m/*z 242 [M] Na^+ was visible.

4.3.1.5 Optimization of kinetic assay

It is possible to monitor the progression of the dehydration by monitoring the consumption of substrate or the formation of product. Typically, the formation of the product is used to track the reaction. Monitoring the formation of the product is more accurate as there is a larger relative change in concentration. The formation of the product can be tracked using the UV peak or the ion mass peak. It was decided that ion mass monitoring would be the most sensitive way to monitor the reaction progression. Monitoring the UV response of the product was unsuitable as the response did not have a linear relationship to concentration, especially at low concentrations (Figure 60).

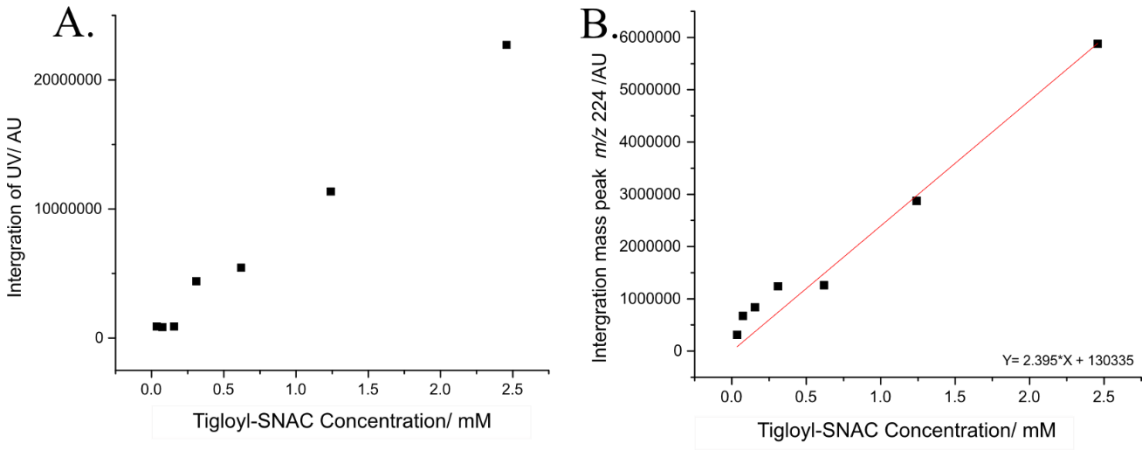


Figure 61- **A**, UV dose response to **186**; **B**, SIM peak area response to **186**.

It was decided that rather than monitoring the formation of product using a total ion count method (TIC), selected ion monitoring (SIM) method would be used. A SIM method was set up to detect the mass of **112**, *m/z* 224 $[M]Na^+$. Optimisation of mass monitoring method from TIC to SIM allows maximum mass sensitivity.

4.3.1.6 Initial time course

To assess the suitability of LCMS as a way of monitoring the kinetics of the DH domain an initial time course was performed. This allowed an understanding of the assay time frame to be developed. The assay protocol used by David Ivison was used. Assays were performed in a volume of 160 µL and contained the following:

Component	Volume (µL)	Final concentration
1 M Tris-HCl pH8	10	62.5 mM
50 % Glycerol	40	12.5 %
Substrate mimic	10	~10 mM
DH (10 $mg.mL^{-1}$)	100	164 µM

Table 2 – SQTKS DH assay conditions, as carried out by David Ivison. [95]

The DH (164 µM) was incubated with 2*R*,3*R*-**114** (1.7 mM), and protein storage buffer pH 8 at 30 °C. Samples of 20 µl were taken after 0.5, 1.0, 2.0, 4.0 and 6.0 hour time points. The aliquots were quenched with MeCN:H_2O (60:20 µl). The precipitated protein was removed *via* centrifugation and the supernatant was then analysed by LCMS (Figure 62). The UV peak of the substrate (2*R*,3*R*)-SNAC **114** decreased over the course of the reaction, while the UV peak associated with the product tiglyl-SNAC increased over the course of the reaction.

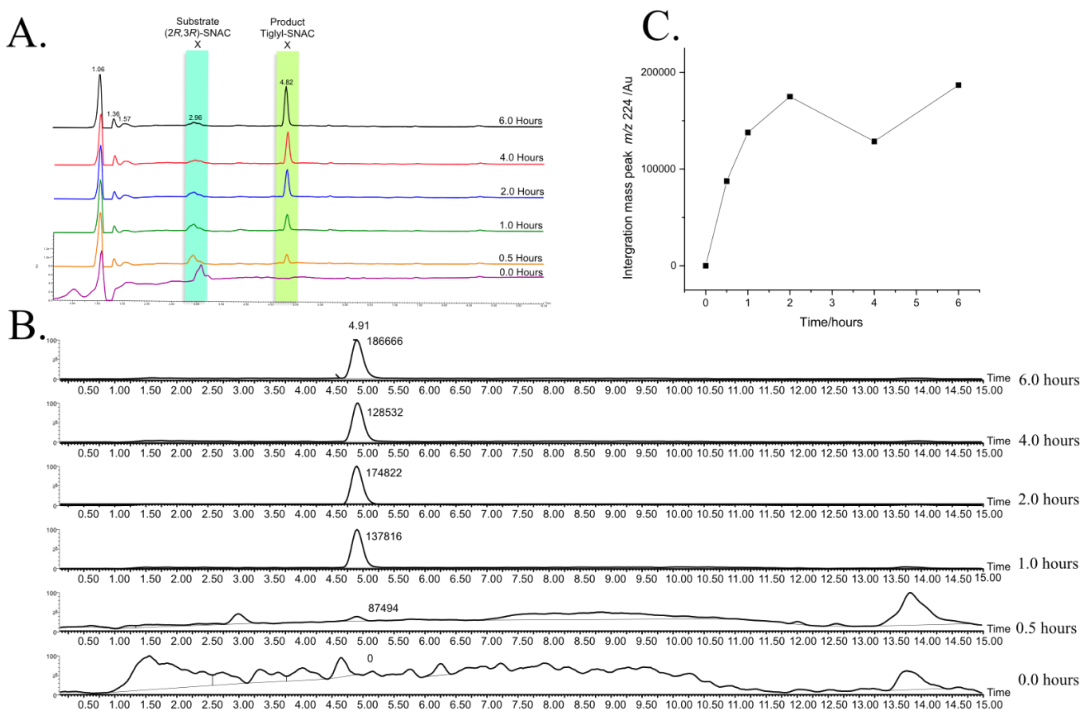


Figure 62 – **A**. Overlay of LCMS UV spectra during time course assay, the increase in the **112** UV peak can clearly be seen over time; **B**. SIM^+ trace of **112**, *m/z* 224 [M] Na^+; C. Graph showing change in SIM^+ peak area of **112** *m/z* 224 [M] Na^+ over time in the presence of DH enzyme.

The peak of area of single ion mass monitoring (SIM) of both the substrate and the product were used to track the progress of the reaction accurately (Figure 62, B). Over time the peak area of 2*R*,3*R*-SNAC **114** decreased and the peak area of tigloyl-SNAC **112** increased. During the initial hour the rate of reaction is linear, after this time the rate of reaction begins to slow. After 3 to 4 hours the change in peak area plateaus and appears to progress no further. The reaching of a plateau is in keeping with findings of other DH assays. It has been found in other DH domains that the dehydration is a reversible reaction and after a certain length of time an equilibrium is reached.[12, 13] A plateau is reached in assays of isolated DH domains due the lack of downstream PKS domains or modules. In the biosynthetic pathway the downstream modules in the PKS are responsible for driving the equilibrium towards to formation of the dehydrated product.[67]

From this it was decided that all subsequent assays would be carried out on a 1 hour time scale as the formation of the product tigloyl-SNAC **112** is linear during this time frame. The effect of protein concentration on the rate of conversion was investigated in another series of time course assays. The initial protocol for the DH assay utilised a high concentration of enzyme (132 µM). Continued use of this concentration of protein is unsustainable due to the time consuming nature of protein production. Therefore it was a priority to find the minimum amount of protein needed to measure the substrate turnover in a realistic time frame. Assays were carried out with various enzyme concentrations in order to establish the limit of detection of the product by SIM-LCMS analysis.

The turnover of substrate 2*R*,3*R*-**114** was measured by SIM LCMS analysis of samples taken at various time points over 6 hours. Assay protein concentrations ranged from 76.9 µM to 60 nM. For assays with a protein concentration less than 5.0 µM, it was not possible to accurately measure substrate turnover within the first hour by LCMS due to the detection limits.

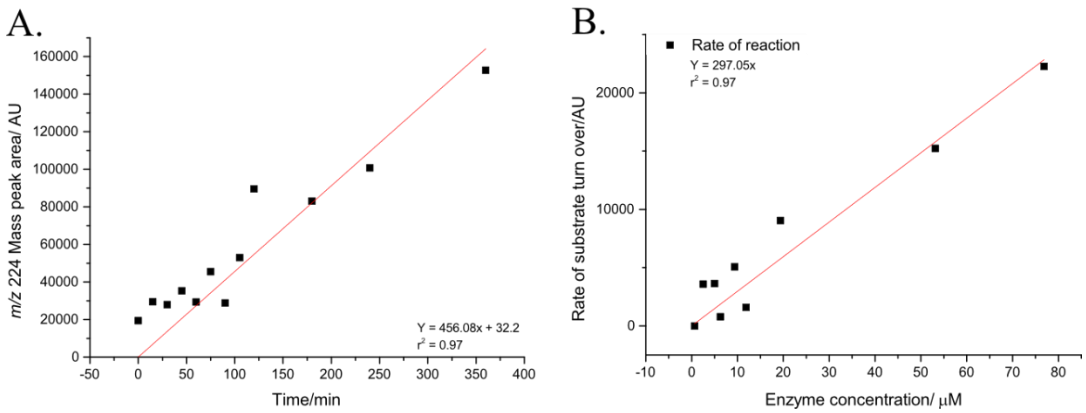


Figure 63 – **A,** Example graph of rate determination at a protein concentration of 9.38 µM; **B**, Rate of reaction plotted against enzyme concentration.

4.3.1.7 Optimization of assay procedure

The results from the enzyme dilution assays showed it was possible to use lower concentration of protein and still measure substrate turnover. It was also decided to modify the assay mixture composition to limit the volume of protein used. As assays were performed over an hour with a maximum of 5 time points, the volume could be reduced from 160 µL to 100 µL. Therefore the assay protein concentration was reduced from 132 µM to 64 µM.

Literature procedures typically required the assay to be quenched, and extracted.[60, 67] The extracts were then dried and redissolved in the correct solvent for MS analysis.

In an effort to simplify the sample processing even further the following protocol was developed. At a certain time point, 20 µL of the assay mixture was taken and the reaction was quenched with MeCN:H_2O (60 µL):(20 µL). The precipitated protein was removed *via* centrifugation and the supernatant transferred to a fresh LCMS vial. The protocol allows samples to be processed with minimal manipulation and also minimizes interfering compounds while allowing protein to be removed.

4.3.2 DH substrate selectivity

4.3.2.1 DH domain substrate specificity

In order to ascertain which of the natural substrate mimics **114**, **115**, **116** and **117** SQTKS DH would be able to act upon, the following assays were undertaken. SQTKS DH (124 µM) was incubated with substrate mimics **114**, **115**, **116** or **117** (0.91 mM) in Tris-HCl buffer (20 mM, pH 8) for 2 hours. Samples of 20 µl were taken every 10 minutes for an hour and then every 30

minutes up to 2 hours. The reaction samples were quenched with MeCN:H_2O (60:20 µl) and the precipitated protein was removed *via* centrifugation and the supernatant was then analysed by SIM LCMS.

The reactions of **115**, **116** and **117** did not show any appreciable increase in the peak associated with the product **112**, *m/z* 224 [M] Na^+, nor did there appear to be any consumption of the substrate ion peak *m/z* 242 [M] Na^+, (Figure 64). In contrast, 2*R*,3*R*-**114** showed substantial turnover of the substrate with a large increase in the 224^+ = [M] Na^+, peak associated with formation of the tigloyl-**112** product (Figure 65). These results indicate that the SQTKS DH domain is highly selective towards substrate with the 2*R*,3*R*-**114** stereochemistry, and is unable to tolerate and turn over any other enantiomer or diastereomer.

Figure 65

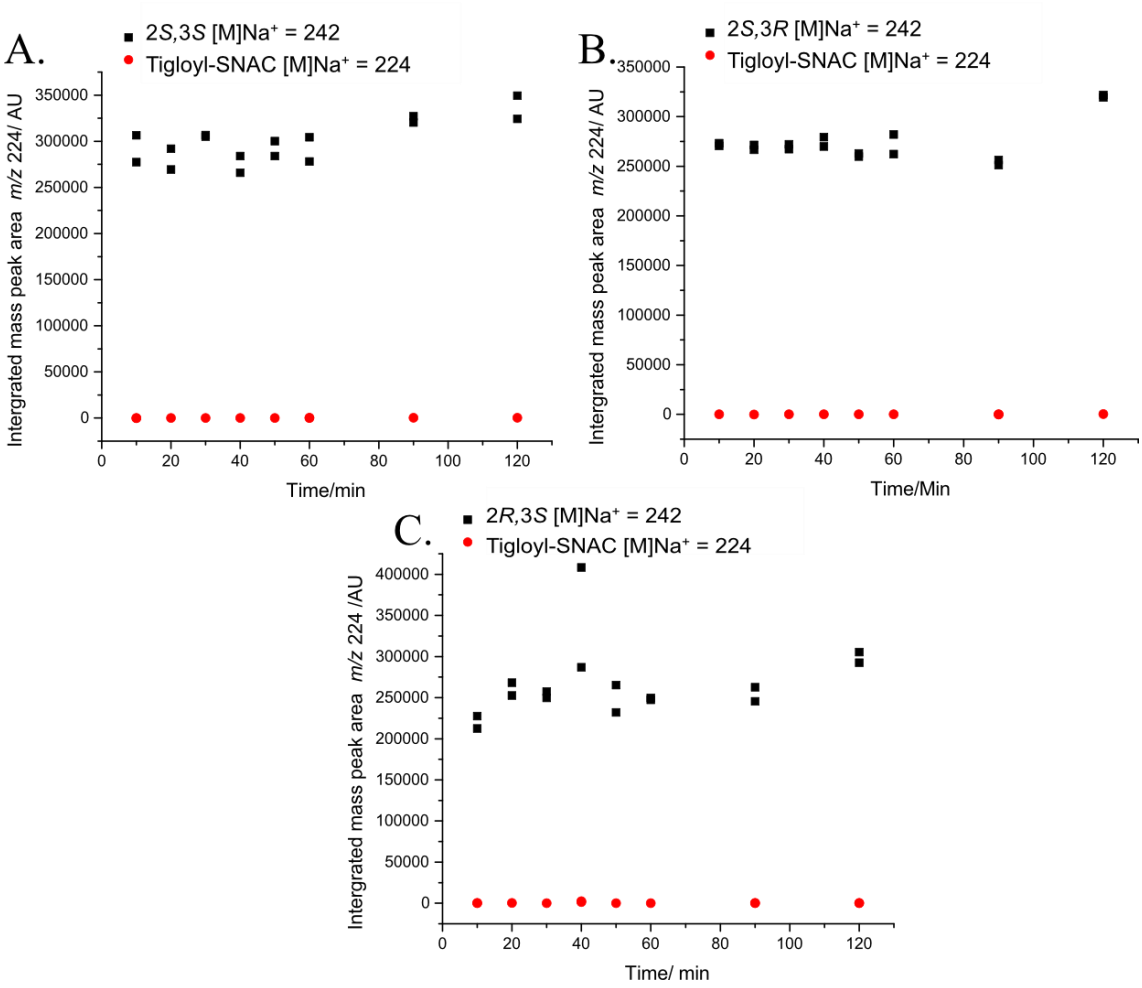


Figure 64 – **A**, Incubation of 2*S*,3*S*-**115** (0.91 mM) with SQTKS DH (132 µM), measuring the integrated peak area of **112** *m/z* 224 [M] Na^+; **B**, Incubation of 2*S*,3*R*-**117** (0.91 mM) with SQTKS DH, measuring the integrated peak area of **112** *m/z* 224 [M] Na^+; **C**, Incubation of 2*R*,3*S*-**116** (0.91 mM) with SQTKS DH, measuring the integrated peak area of **112** *m/z* 224 [M] Na^+.

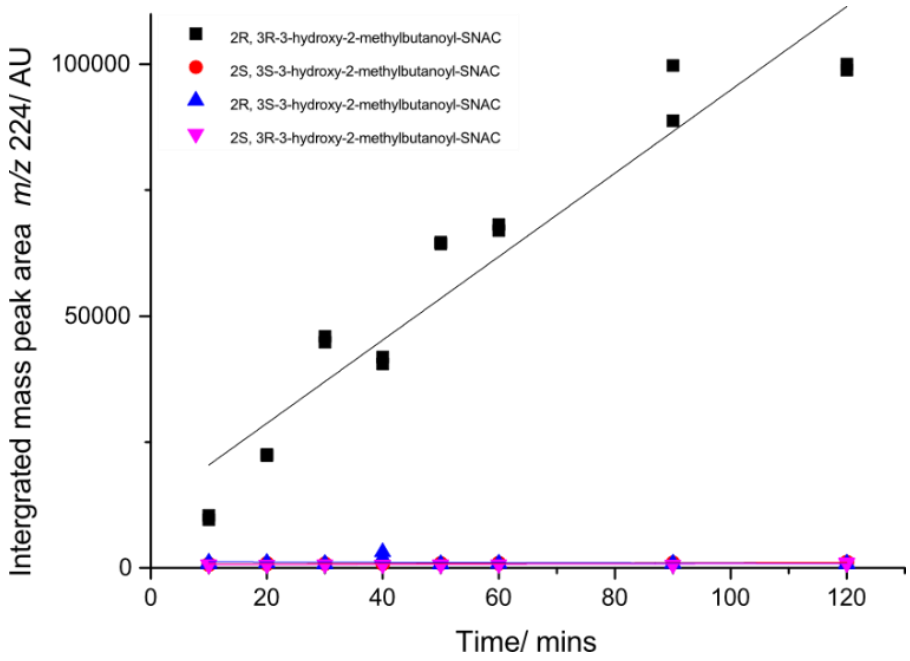


Figure 65 – Incubation of various substrates **114**, **115**, **116** and **117** (0.91 mM) with SQTKS DH (132 µM),peak area of product tigloyl-SNAC **112** , ion mass 224 [M] Na^+.

In order to probe the DH active site further, the isolated DH protein (132 µM) was incubated with non-natural substrates **124** and **125** (0.97 mM) in Tris-HCl buffer (20 mM, pH 8) for 2 hours. Samples of 20 µl were taken every 10 minutes for an hour and then every 30 minutes up to 2 hours. The reaction samples were quenched with MeCN:H_2O (60:20 µl), the precipitated protein was removed *via* centrifugation and the supernatant analysed by SIM LCMS.

The non-natural diastereomeric substrate mimics **124** and **125** were designed to assess the importance of the α-methyl group for correct positioning in the active site. Both substrates lacked the α-methyl group but retained the β-hydroxyl group. The reaction of (3*S*)-**125** did not show any significant increase in the size of the peak associated with the product crotonoyl-SNAC **185** *m/z* 210 [M] Na^+, (**Figure 66**). The reaction of (3*R*)-**124** showed the substrate underwent dehydration with a significant increase in the SIM peak associated with the product 1**85** *m/z* 210 [M] Na^+ (Figure 66, Figure 67).

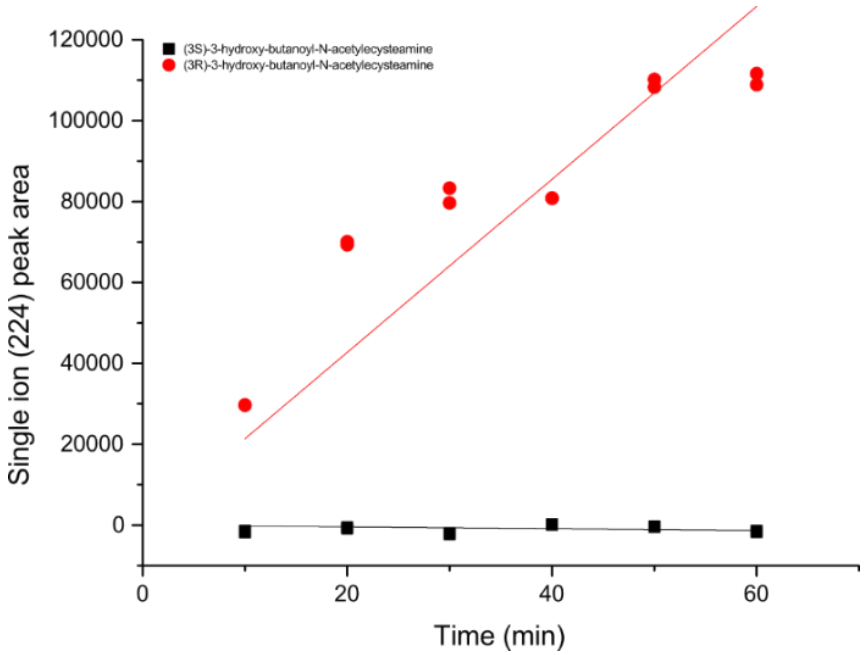


Figure 66 - Incubation of substrates **124** and **125** (0.97 mM) with SQTKS DH (132 µM), peak area of product **185**, ion mass 210 [M] Na^+.

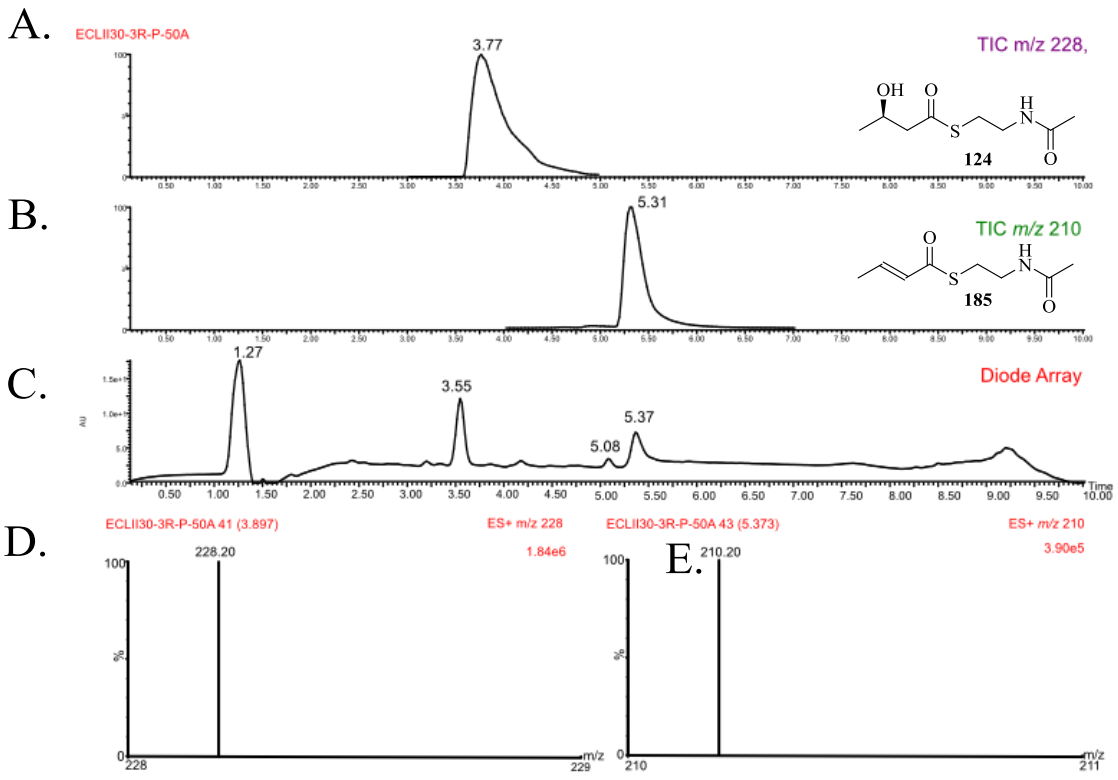


Figure 67 - **A**. ESI+ SIM trace of 3*R*-**124**; **B.** ESI+ SIM trace of **185; C**. UV trace of 3*R*-**124** assay; **D**. ESI+ of substrate **124** [M]Na^+ = 228**; E,** ESI+ of product **186** [M] Na^+ = 210;

4.3.3 Inhibition studies

The investigation into the stereoselectivity of the DH domain found that the DH is highly selective with only 2 of the substrate mimics tested being competent, these are 2*R*,3*R*-**114** and 3*R*-**124**. In order to gain more insight into the programming of the active site, inhibition assays were undertaken.

The aim of these assays was to assess why the DH domain is unable to turn over the non-competent substrates. There are two possibilities as to why the DH domain is unable to catalyse the dehydration of the non-competent substrates. It is possible that the non-active substrate are able to enter the active site but the catalytic dyad is unable to recognise the key 2-hydrogen and 3-hydroxy groups to facilitate the reaction. If this is the case it would be expected the compounds could act as inhibitors. It is also possible the non-active substrates were not able to enter the active site at all. In this case the compounds should not be inhibitors of the DH.

The SQTKS DH (52.5 μM) was incubated with mixtures of 2*R*,3*R*-**114** (2.28 mM) and either **115**, **116** or **117** (2.2 mM) in Tris-HCl buffer (20 mM, pH 8) at 30 °C for 40 minutes. Samples of 20 μl were taken at 5, 10, 20, 30 and 40 minute time points. The reactions were quenched with MeCN:H_2O (60:20 μl), the precipitated protein was removed *via* centrifugation and the supernatant analysed by SIM LCMS. Control assays were carried out in which the conditions were not altered apart from **115**, **116** and **117** being replaced by distilled water, this allowed the uninhibited rate of substrate turnover to be monitored in parallel and the two rates compared.

The rate of turnover of 2*R*,3*R*-**114** to the tigloyl-product was not inhibited in any of the non-competent substrates **115**, **116** or **117** (Figure 68). The findings indicate that the non-active substrates do not enter the active site of the DH domain as the rate of turnover of 2*R*,3*R*-**114** is not inhibited at all. If the non-competent substrates were able to occupy the active site for any length of time, the active substrate 2*R*,3*R*-**114** would be unable to enter the active site and the rate at which tigloyl-**112** is formed would decrease.

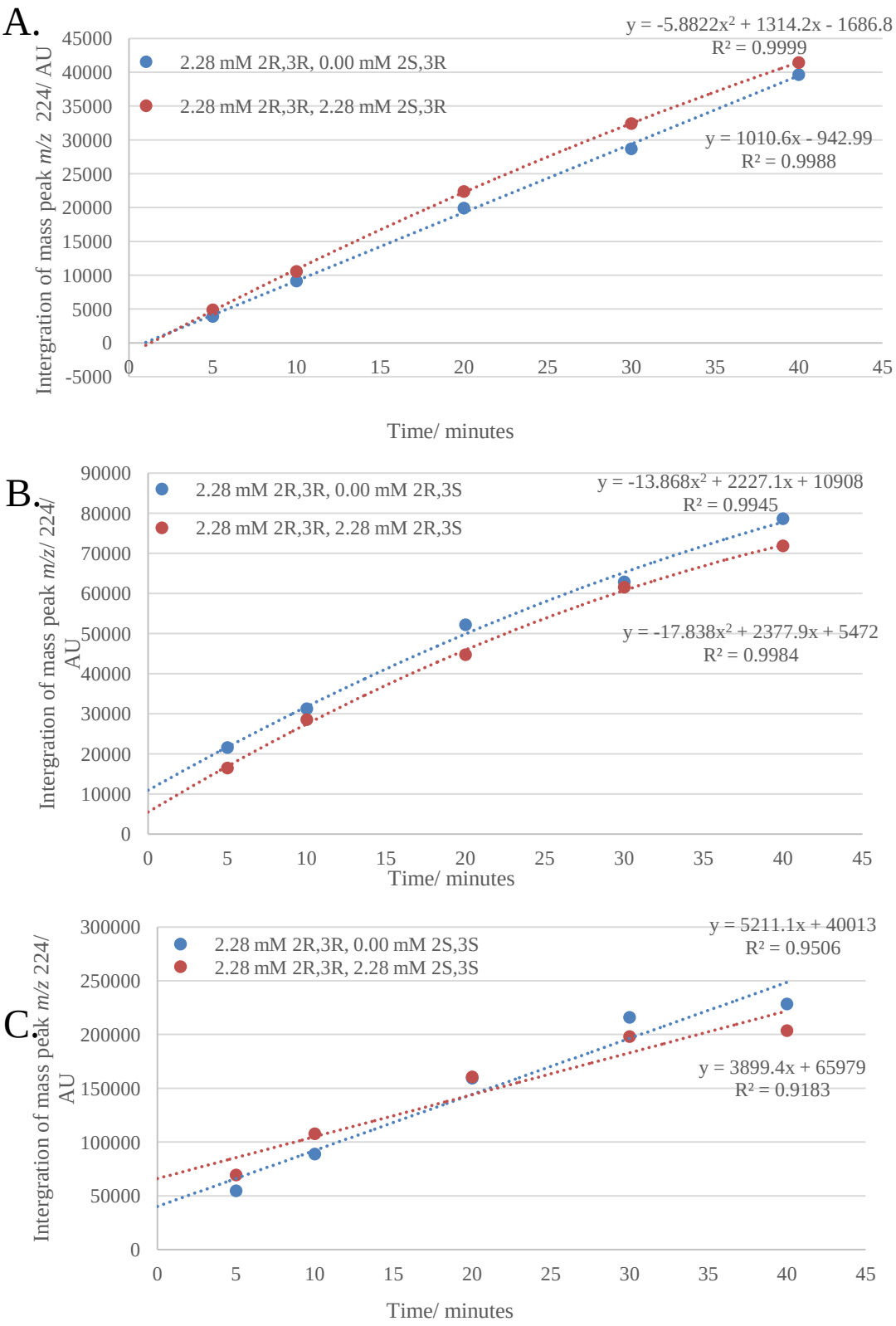


Figure 68 – **A**, Inhibition studies of active substrate 2*R*,3*R*-**114** vs 2*S*,3*R*-**117** (2.28 mM) with SQTKS DH (132 µM), peak area of product **112**, ion mass 224 [M] Na^+; **B**, Inhibition studies of active substrate 2*R*,3*R*-**114** vs *2R*,3*S*-**116** (2.28 mM) with SQTKS DH (132 µM), peak area of product **112**, ion mass 224 [M] Na^+; **C**, Inhibition studies of active substrate 2*R*,3*R*-**114** vs 2*S*,3*S*-**115** (2.28 mM) with SQTKS DH (132 µM),peak area of product **112**, ion mass 224 [M] Na^+.

4.3.4 Kinetic studies of SQTKS DH

4.3.4.1 Kinetic studies of substrate 2*R*,3*R* -114

In order to assess the binding affinity of the 2*R*,3*R*-**114** substrate towards the SQTKS DH and how effective the DH is at catalysing the dehydration reaction, efforts were made to measure the Michaelis-Menton parameters. The initial rate at which the isolated DH domain catalysed the dehydration of 2*R*,3*R*-**114** was measured at various concentrations ranging from 0.23 mM to 4.5 mM. The SQTKS DH (60 µM) was incubated with 2*R*,3*R*-**114** in Tris-HCl buffer (20 mM, pH 8) at 30 °C for 40 minutes. Samples of 20 µl were taken at 5, 10, 20, 30 and 40 minute time points. The reactions were quenched with MeCN :H_2O (60:20 µl), the precipitated protein was removed *via* centrifugation and the supernatant analysed by SIM LCMS. The assays were repeated in triplicate. Control assays were also carried out in tandem, using identical conditions where the protein was replaced with purification buffer. The initial rate of dehydration plots for **114** are shown in Appendix 2.

As the concentration of the substrate **114** decreased, the initial rate of dehydration also decreased as expected (

Table **3**). The initial rate for each concentration was calculated and plotted into a Michaelis-Menton plot (Figure 69). In Michaelis-Menton kinetics, the initial rate v_0 increases linearly with concentration till V_{max} is approached, where v_0 plateaus. The data shows a linear increase in v_0 with increasing concentration. The data from the two highest concentrations (3.0 mM and 4.5 mM) may indicate the initial stages of v_0 plateauing, without additional data for higher concentrations this remains unclear.

[S] (mM)	$\mathbf{V_o}$ (µM min^{-1})
4.5	5.40
3	3.78
2.3	2.74
1	3.00
0.8	1.46
0.5	1.16
0.2	0.80

Table 3 - Velocity of SQTKS DH enzyme catalysed dehydration reaction at various concentrations of substrate 2*R*,3*R*-**114**.

While it is not possible to calculate the $V_{max,}$ and therefore calculate the k_{cat} or K_M the from the data obtained, it is possible to calculate the substrate efficiency, k_{cat}/K_M by measuring the gradient at the origin of the Michaelis-Menton plot (Figure 69). The gradient at the origin is calculated by differentiating the equation of the curve fitted on the Michaelis-Menton plot

(Figure 69 and Eq6 - 8). This results in a general expression of the gradient. The k_{cat}/K_M for the 2*R*,3*R*-**114** substrate is 1.9 min^{-1} mM^{-1}.

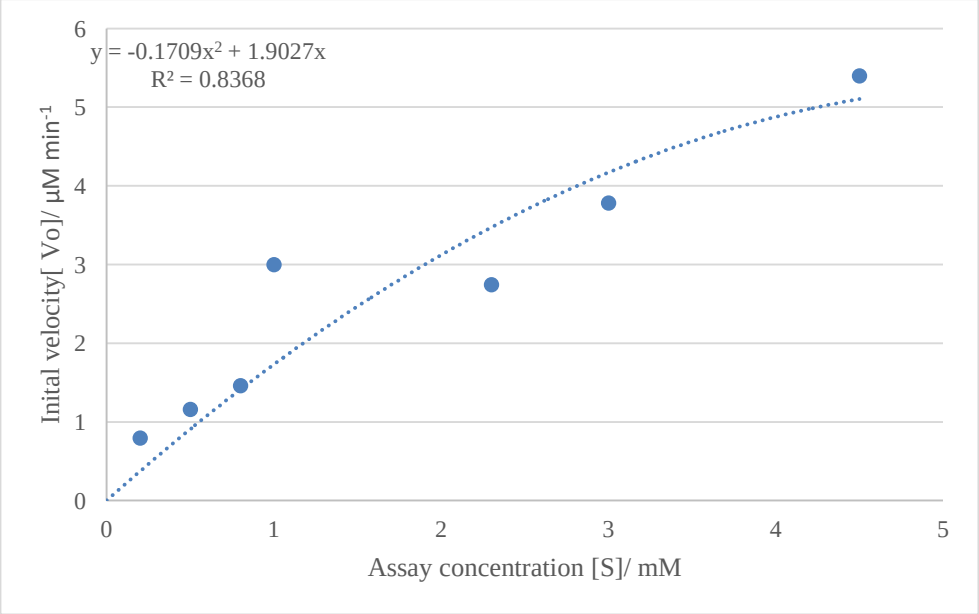


Figure 69 – Velocity of SQTKS DH enzyme catalysed dehydration reaction as a function of substrate concentration of 2*R*,3*R*-**114**.

$$y = ax^2 + bx + c \quad \text{Eq. 6}$$

$$\frac{dy}{dx} = 2ax + b \quad \text{Eq. 7}$$

$$if\ x = 0, then\ \frac{dy}{dx} = b \quad \text{Eq. 8}$$

4.3.4.2 Kinetic studies of substrate 3*R* -124

An investigation to assess the binding affinity of the non-natural substrate 3*R*-**124** was also undertaken. The initial rate of dehydration of 3*R*-**124** was measured over a range of concentrations (0.2 mm – 3.99 mM), initial rate plots are shown in Appendix 3. The assay conditions used were identical to those used in previous kinetic studies. The data showed that the initial rate of dehydration increased as the concentration of 3*R*-**124** increased (Table 4). The initial rates appeared to plateau, the calculated initial rates were plotted into a Michaelis-Menton plot (Figure 70).

[S] (mM)	**V_o** (µM min^{-1})
3.99	1.06
3.5	0.90
2.6	1.11
2	0.68
0.88	0.43
0.64	0.35
0.44	0.13

0.2	0.18

Table 4 - Velocity of SQTKS DH enzyme catalysed dehydration reaction at various concentrations of substrate 3*R*-**124**.

The Michaelis-Menton plot showed a linear relationship between the initial rate and concentration of 3*R*-**124**. At higher concentrations, 2.6 mM to 3.5 mM the initial rate appears to plateau although without additional data this can not be confirmed. It is therefore, as with the 2*R*,3*R*-**114** substrate, not possible to calculate $V_{max,}$ k_{cat} or K_M the from the data obtained. The substrate efficiency, k_{cat}/K_M, was calculated in the same way as the 2*R*,3*R*-**114** in Section 4.3.4.1, and found to be 0.5 min^{-1} mM^{-1}.

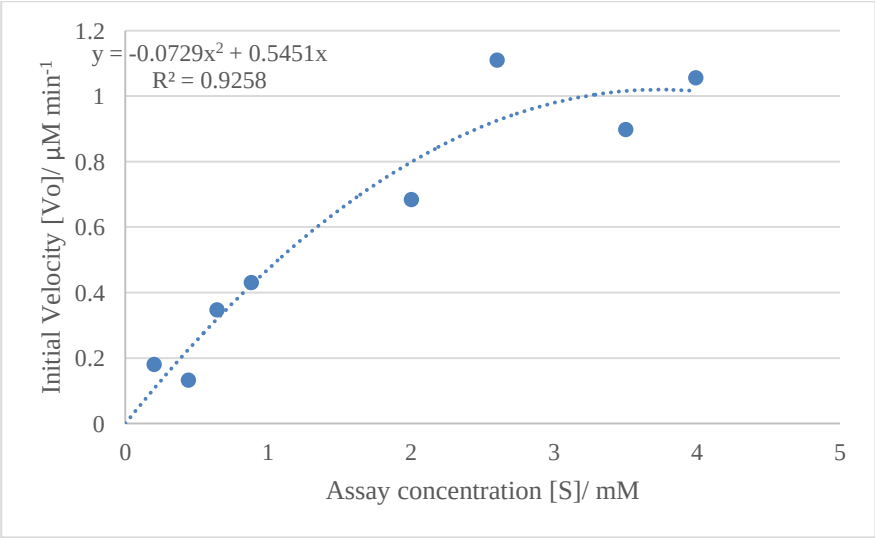


Figure 70 - Velocity of SQTKS DH enzyme catalysed dehydration reaction as a function of substrate concentration of 3*R*-**124**.

4.4 Conclusion

The isolated SQTKS DH domain was found to be highly selective, only catalysing the dehydration of two substrates, one natural substrate mimic, 2*R*,3*R*-**114** and one non-natural, 3*R*-**124**. This finding indicates that the DH domain active site is highly programmed to only recognise a substrate with a 3*R*-hydroxyl group. The active site also only tolerates a non-methylated substrate or substrate with a 2*R*-methyl group. Inhibition studies were carried out to assess the effect of non-active substrates **115**, **116** and **117** on the initial rate of dehydration of 2*R*,3*R*-**114**. All three substrates were found to not inhibit the rate of dehydration of 2*R*,3*R*-**114**. This indicates that the SQTKS DH domain is programmed to only allow certain substrates into the active site.

The stereochemistry of the non-natural substrates shows that the 3-hydroxy position is key to active turnover. Without the correct stereochemistry at the hydroxyl position, dehydration is unable to occur. It is likely that the programming selectivity is the result of the positioning of the catalytic amino acids within the active site. This will be discussed further in chapter 5.

Kinetic studies were undertaken to gain more insight into the bind affinity and efficiency of both substrates. The data obtained was not ideal, as it was not possible to calculate $V_{max,}$ k_{cat} or K_M for either substrate due to slow and inefficiency substrate turn over. It was possible to calculate the substrate efficiency k_{cat}/K_M, although these values must be treated with caution as the errors on the initial rate of dehydration kinetic data are large.

The k_{cat}/K_M for 2*R*,3*R*-**114** was calculated to be 1.9 min^{-1} mM^{-1}, while 3*R*-**124** was calculated to be 0.5 min^{-1} mM^{-1}. k_{cat}/K_M describes how effectively a substrate is turned over by the active site, with the smaller k_{cat}/K_M, indicating a substrate is turned over less efficiently. The much lower k_{cat}/K_M for 3*R*-**124** indicates the presence of a methyl group is key to efficient substrate turnover. The impact of stereochemistry and the presence of the methyl group will be examined and discussed in more detail in chapter 5.

The first PKS DH domain to undergo steady-state kinetic characterisation was the pikromycin DH domain, Pik2DH.[67] Aldrich *et al.* utilized LC-MS/MS to monitor the reaction and found the PikDH2 domain to be highly selective. Substrates **179a-b** were designed to test the stereospecificity of the PikDH2 domain exhibited towards the β-stereocenter. Substrates **180a-e** were designed to assess how sensitive PikDH2 was towards changes in stereochemistry at the γ and δ positions. Of the 7 substrates tested, only the substrate with the native stereocenters **179a** and **180a** were accepted and processed by the enzymes active site. The k_{cat}/K_M of **179a** was calculated to be 0.097 min^{-1} mM^{-1}, while **180a** was calculated to be 0.225 min^{-1} mM^{-1}.[67] The k_{cat}/K_M of the SQTKS substrate 3*R*-**124** is similar to that of **179a** and **180a**. All three substrates are non-methylated at the β-positions.

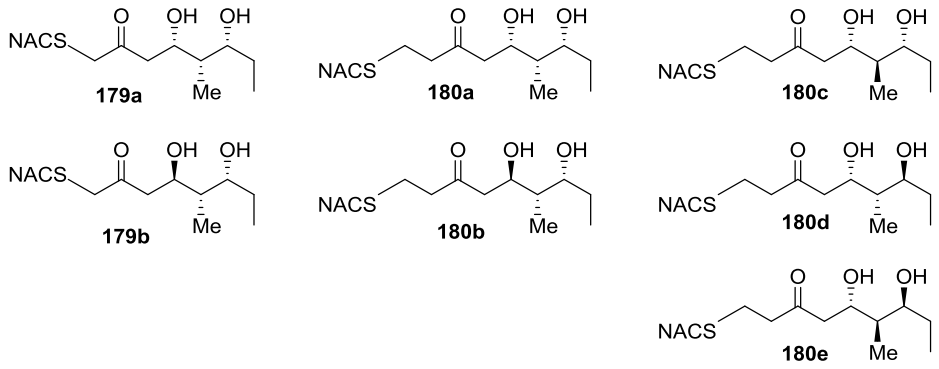


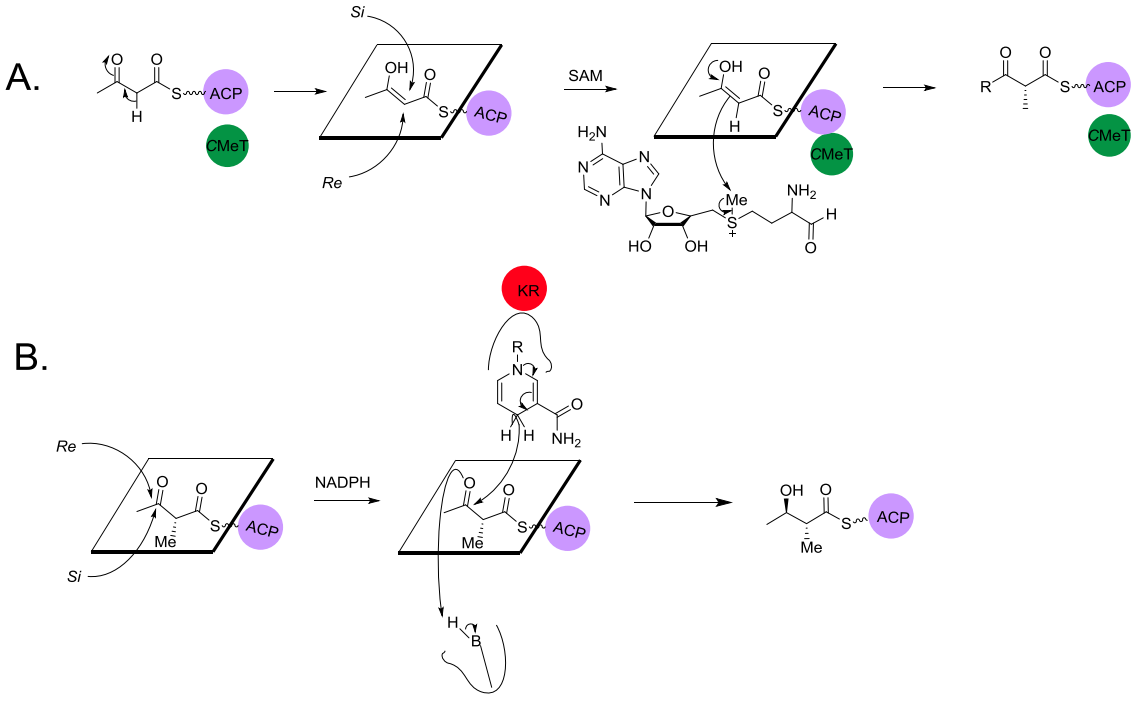
Figure 71 – Structures of substrates used to study the steady-state kinetics of PikDH2.

As the stereochemistry of the final SQTKS polyketide is known, it is therefore known that the SQTKS ER catalyses the reduction of the α,β-unsaturated thiolester giving the methyl group *S* stereochemistry. The stereochemical outcome of the earlier β-keto processing reactions is unknown. The uncovering of the SQTKS DH domains stereoselectivity allows the elucidation of stereochemical outcome of the KR reduction. The stereoselectivity of the DH domain may

also indicate the selectivity of the *C*Met methylation, although it is likely the product of the *C*Met methylation will racemise easily.

The DH diketide substrate 2*R*,3*R*-**114** possesses a 2*R* methyl group. Therefore the *C*Met domain must catalyse the methylation of the α-position so an *R* stereocenter is formed. Assuming there is no racemisation of the *C*Met product, it can be assumed that the programming of the *C*Met domain causes the methylation to occur so the methyl group is added to the *Re* face of the alkene ester, as (Scheme 58,A).

It is also possible to predict the stereochemical outcome of the reduction reaction. The DH domain substrate has been found to be 2*R*,3*R*-**114**, therefore the KR domain must catalyse the reduction of the β-keto to form the 3*R*-hydroxyl group at the β-position. Therefore the reduction must occur with the addition of a hydrogen to the *Si* face of the β-ketone (Scheme 58,B).



Scheme 58 – **A**, *C*Met catalysed methylation of SQTKS diketide with the methylation occurring on the *Re* face of the alkene; **B**, KR catalysed reduction of SQTKS diketide, with the addition of a hydride to the *Si* face the β-keto group.

4.5 Future work

Once the triketide substrate mimic **168** has been successfully synthesised it should be assayed with the isolated DH domain and the results compared with the results from the diketide kinetic assays.

Chapter 5 – Modelling the DH domain

5.1 Introduction

To date it has not been possible to obtain a crystal structure of the SQTKS DH domain (Section 3.4). As a crystal structure is unavailable, efforts were made to create a model which would assist in the explanation of stereo selectivity exhibited by the DH domain. The course of the DH mechanism was discussed in Chapter 1, Section 1.5.1. The structure of crystallised DH domains from FAS (FabA, FabZ, mFAS) and PKS (CurDH F, H, J, K, Ery4DH and Rif10DH) were discussed in detail in chapter 1, Section 1.5.3. The findings from investigations into substrate selectivity of the aforementioned DH domains are also discussed in this section.

5.1.1 Available crystal structures

In order to build a suitable model of the SQTKS DH domain, homologous dehydratases with available crystal structures had to be identified. The RSCB protein data bank (PDB) is a searchable repository of experimentally-derived biological molecules.[144] It is possible to search the database for protein structures which show sequence similarity to SQTKS DH using the NCBI PBLAST algorithm.[76, 144]

A search of the PDB database for protein sequences similar to the SQTKS DH domain found eight homologous sequences: four PKS DH domains, (CurF, CurK, CurJ and CurH)[17] from the curacin pathway; the terminal DH, (Rif10DH),[13] from the rifamycin pathway, Ery4DH[6] from the erythromycin pathway and mycocerosic acid DH domain (MASDH)[145]. The DH domain from the PKS PpsC (PBD: 4OOC) from mycobacterium tuberculosis was also available but was awaiting full publication.[146] The crystal structure of the DH domain from mFAS is also available.[34, 48]

5.1.2 Composition of DH domain

The DH crystal structures from mFAS, curacin, rifamycin, erythromycin and mycocerosic acid were discussed in Sections, 1.5.4, 1.5.5, 1.5.6, 1.5.7 and 1.5.8. Five key structural features were observed in all the available crystal structures: a double hotdog fold; an active site comprised of a histidine and an aspartic acid; a conserved proline situated near the catalytic histidine; a hydrogen bonding network; and a substrate tunnel (Figure 72).

The double hotdog fold, where parallel β-sheets wrap around two hydrophobic α-helices, is a highly conserved structural feature in all DH domains. It is key in shaping the internal structure and contributing to the formation of the substrate tunnel. The substrate tunnel opening forms between the two α-helices and the substrate tunnel stretches through the C-terminal hotdog fold.

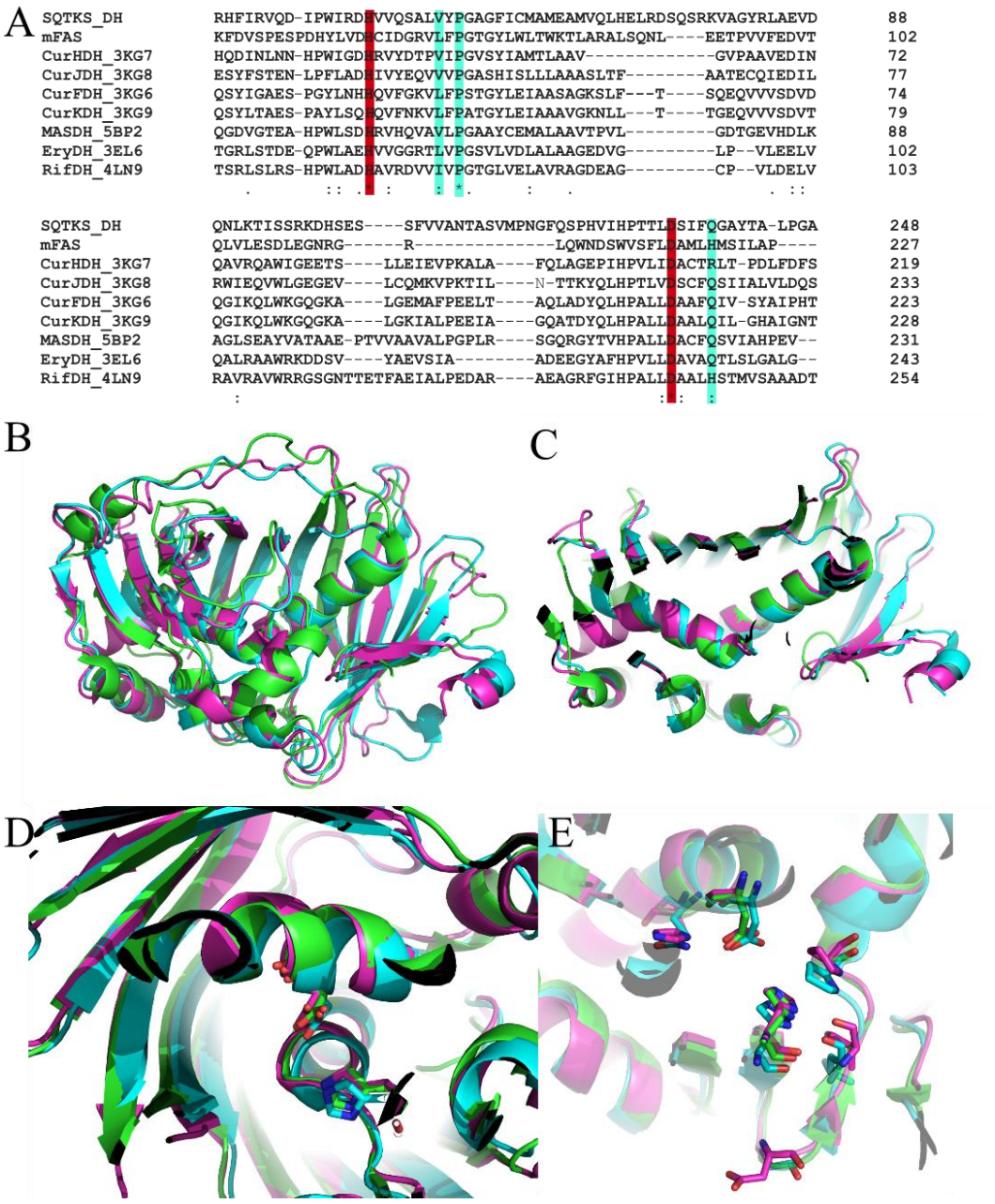


Figure 72 – **A**, Key Sections of sequence alignment, showing catalytic amino acids histidine and aspartic acid shown in red. Amino acids involved in the hydrogen bonding network are labelled in blue; **B-E**, Overlay of crystal structures of CurFDH (Green), EryDH4 (Blue) and RifDH10 (purple); **B,** overall structure of the three DH domains; **C**, hotdog folds in all three DH domains; **D,** catalytic amino acids of all three domains; **E,** hydrogen bonding network in all three DH domains.

The substrate tunnel is hydrophobic, generally lacking in hydrogen bonding groups and with a shape which is largely untailored to the exact shape of the substrate.[6] The catalytic amino acids are positioned with the histidine on the N-terminal hotdog fold and the aspartic acid residue residing on the C-terminal hotdog fold.[6, 17] In addition to the catalytic amino acids, a number of other amino acid residues are conserved and form a hydrogen bonding network within the active site.[6] The purpose of this hydrogen bonding network is to ensure the correct positioning of the catalytic amino acids. The proline in the HXXXGXXXXP motif is highly conserved in both FAS and PKS DH domains (Figure 72, A). The proline helps orientate the catalytic histidine

through Van der Waals interactions. Hydrogen bonding from the backbone carbonyl of a nearly invariant leucine (sometimes valine, Figure 72, A) also contributes to the correct orientation of the histidine. It is also thought that the hydrogen bonding network contributes to the pK_a of the aspartic acid residue ensuring that it is able to donate a proton to the substrate, facilitating the subsequent loss of water.[6, 13]

5.2 Model building

5.2.1 Identifying a suitable candidate

It is possible to create a model of the DH domain by *threading* the sequence of the SQTKS DH domain on to the structures of proteins which are available and then minimising the side chain steric and electronic clashes. By selecting protein structures which catalyse a similar reaction and are homologous in sequence, the model formed should reflect the true structure of the DH domain.

A search of the PDB identified published crystal structures was carried out, in order to assess their sequence identity to SQTKS DH and therefore their suitability to take forward to build a model. Sequence alignment was carried out using Clustal Omega (Appendix 1).[76, 77, 147] In addition to aligning the sequences of all the available DH domains, Clustal Omega was able to calculate their sequence identity to the SQTKS DH domain (Table 5).

	% sequence identity to SQTKS
SQTKS	100.0
4LN9 Rif10	18.0
3EL6 Ery4	16.9
3KG7 CurH	17.3
3KG8 CurJ	18.1
3KG6 CurF	19.4
3KG9 CurK	18.2
5BP2 MASDH	25.5

Table 5 – Crystallised DH domains % sequence identity to SQTKS DH domain.

Rif10DH, Ery4DH, CurJ, CurK and CurH all had similar identities ranging from 16.9% to 18.2%. Curacin CurF DH had the greatest sequence identity of all the curacin DH structures at 19.4%. MASDH had a sequence identity of 25.5%, the highest identity of all the available crystal structures. Although the sequence identity for all the DH domains is low, it had been found that despite this, they have a high degree of structural similarity. Akey *et al.* compared the structural and sequence identity of CurFDH with the other CurDH domains, Ery4DH.[17] While CurK had a sequence identity to CurFDH of 68%, the other domains had a much lower sequence identity, ranging from 20-32%. The root-mean-square deviation of atomic positions (RMSD) measures the average distances between the atoms of two superimposed proteins.

When the structures were aligned and the RMSD of the α-Carbon was compared, they were found to have a high structural identity with RMSD ranging from 1.0-2.0Å.

5.2.2 Modelling programs

Typically computational modelling programs produce a 3D protein model in one of three ways. The programs can construct the model using comparative modelling (CM), where the sequence or sequence profile comparison identifies evolutionarily related homologous templates.[148, 149] The model is then generated by simply copying the framework from the template. This strategy results in the generation of a high resolution model.

Modelling programs which use a threading method match the query sequence directly onto solved 3D protein structures by recognising similar folds in the query sequence. This method identifies proteins which have similar structures but have different evolutionary origins. For protein sequences which have no structurally related proteins in the PBD, *ab initio* (or de novo) modelling must be carried out. This technique predicts the tertiary structure on the amino acid sequence of the query protein. The success of this type of modelling is limited to small proteins only.

A range of free to use online modelling programs are available, such as RaptorX,[150] HHpred,[151], Robetta[152] and SWISS-MODEL.[153] Ideally, in order to get an accurate model of SQTKS, solved crystal structures of proteins with a high sequence homology would be available. Unfortunately, there are only eight published crystal structures of isolated PKS DH domains, all with low sequence identity to SQTKS, ranging from 16.9% to 25.5% In order to try and overcome this issue, a number of models were produced using different modelling programs. Different modelling programs will create the models using different algorithms and databases. The following modelling programs were used: SWISS-MODEL, Robetta, Phyre 2, Raptor X, HH pred and Itasser.

5.2.3 Building the protein models

5.2.3.1 Itasser[148, 154]

I-TASSER uses a composite approach to building protein models, utilising CM, threading and de novo modelling. From the submitted amino acid sequence, I-TASSER generates three-dimensional models from multiple threading alignments and structural assembly simulations. The output consists of a predicted secondary structure and five 3D models (Figure 73). Each model is assigned a *C*-score, which estimates the quality of the predicted model. *C*-score values typically range from -5 to 2, a higher *C*-score, indicates a higher quality of model. Models with a *C*-score of -1.5 or greater are generally thought to have a correct fold. Of the five models created by the I-TASSER modelling server only two had *C*-scores higher than -1.5, those were

models 1 and 4. I-TASSER model 1 had the highest *C*-score of 0.56, therefore only model 1 was taken forward to compare with models produced by other modelling programs.

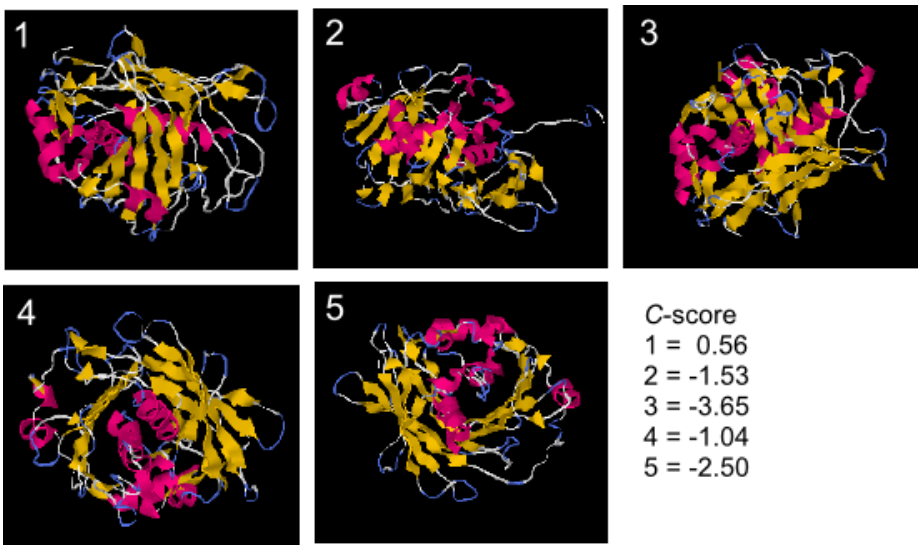


Figure 73 – I-TASSER models of the SQTKS DH domain and associated *C*-scores.

5.2.3.2 Phyre2[155]

The Phyre2 server offers a simple and intuitive portal for protein modelling.[155] It is possible to submit your model to be processed in two modes, normal and intensive. The intensive mode creates a full length model using multiple template modelling and simplified *ab initio* folding simulations for sections of the sequence not covered by the template. The SQTKS DH domain sequence was submitted to the intensive modelling server. Phyre2 selected 6 templates, Rif2DH and the 4 Curacin DH domains; CurF, CurH, CurJ and CurK. It also selected the dehydratase domain of rzxb from Pseudomonas Fluorescens. Phyre2 modelled 95% of the SQTKS protein at a confidence of >95%. The server was not able to model 16 residues using the selected templates and had to model them *ab initio*. It must be noted that *ab initio* modelling is highly unreliable. Overall the templates selected by the server covered the majority of the SQTKS sequence.

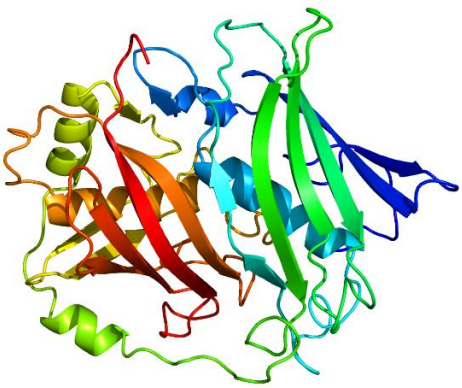


Figure 74 – Phyre2 model of the SQTKS DH domain.[155]

5.2.3.3 RaptorX[150, 156]

RaptorX uses template based modelling (TBM) to generate a protein model.[150, 156] RaptorX is able to carry out three different modelling tasks, tertiary structure modelling, secondary structure modelling and custom alignment modelling. The amino acid sequence for SQTKS DH was submitted to the RaptorX for secondary structure modelling and resulted in one protein model (Figure 75). RaptorX used the curacin DH domain CurJ (3Kg8) as a template to thread the SQTKS-DH sequence. The protein model produced by RaptorX is given an unnormalized global distance test (uGDT) score, which measures the absolute model quality. For proteins of more than 100 residues a uGDT score of 50 or more is an indicator of a good model. The SQTKSDH model contains 324 residues and was given a uGDT of 183, indicating the SQTKS-DH model generated by RaptorX is acceptable. The model also gave a global distance test (GDT) score. A poor GDT score is defined as a score less than 50. A good uGDT score but a poor GDT score is indicative of a model in which only a small portion of the model is accurate. The GDT score given to the RaptorX SQTKS DH model was 56 which is lower than desired and may indicate that errors are present in this model.

Figure 75 – RaptorX model of the SQTKS DH domain

5.2.3.4 Robetta[157-159]

Robetta provides an automated structure prediction service which aims to create a protein model from genomics data. The initial step of the modelling process uses a domain prediction method called ‘ginzu’. Ginzu is a hierarchical screening process which finds regions in the query sequence which are homologous to experimentally determined structures. Regions which are homologous to solved structures undergo comparative modelling. Any regions which are not covered undergo *de novo* modelling. The amino acid sequence for SQTKS was submitted to Robetta for structure prediction and resulted in 5 models. Models 1 and 4 were produced by comparative modelling, the DH from mycocerosic acid synthase-like (MASDH) PKS was used as the template for these models. Models 2, 3 and 5 were produced by *de novo* modelling. This

form of modelling is known to be inaccurate and is typically only used for small proteins (less than 150 amino acids). Therefore models 2, 3 and 5 were discounted and no further analysis was undertaken using these models. Models 1 and 4 were aligned in PyMol and found to be similar, although there were some differences in the predicted positioning of some α-helices and β-sheets (Figure 76, F).

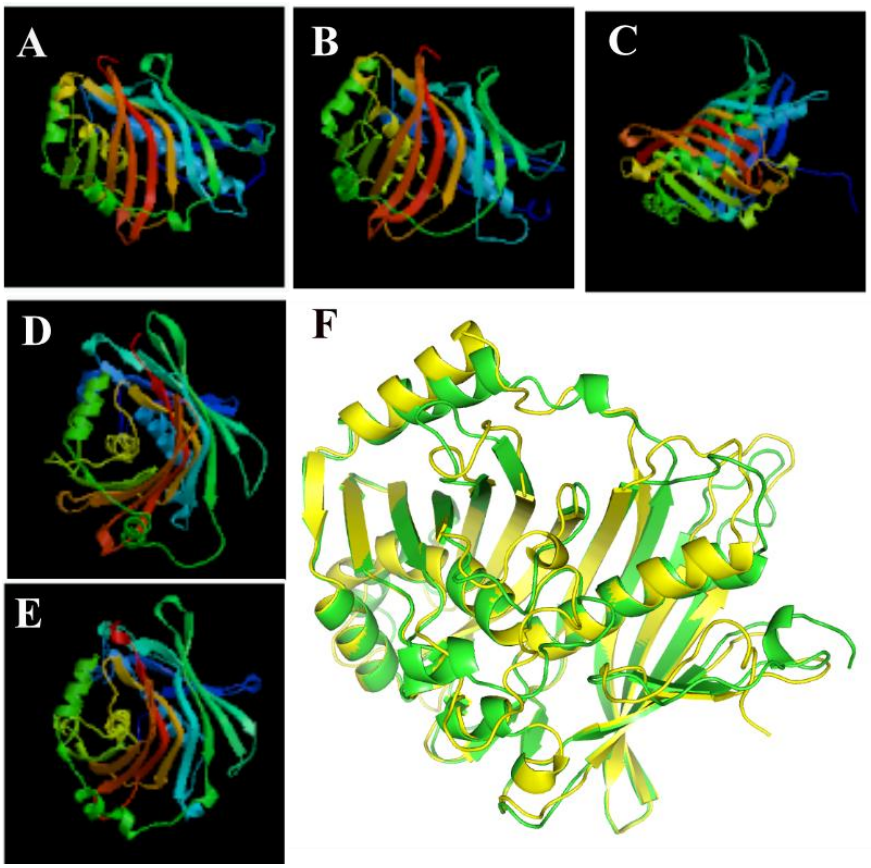


Figure 76 – A-E; Robetta models of the SQTKS DH domain, A, model ; B, model 2; C, model 3; D, model 4; E, model 5; F, PyMol alignment of models 1 and 4.

5.2.3.5 HH Pred[160]

HHPred is a remote server which can be used to quickly assess protein homology and carry out structure prediction. The SQTKS sequence was submitted to the server as a single query and searched against the PDB database. HHPred returns multiple alignments of the query in the form of a "hit list" of similar sequences and ranks them according to probability. The results also contain details of the alignment of each result against the query sequence with annotations on secondary structure. The highest scoring sequences were MASDH (99.75), CurJ (99.44), CurH (99.39) and CurK (99.32). HHPred uses the alignment results to automatically select the most suitable proteins and the automatic selection tool chose the following protein structures to produce the template, MASDH, CurJDH, CurK, CurF. The template was subsequently submitted to Modeller and produced one model (Figure 77).[161, 151]

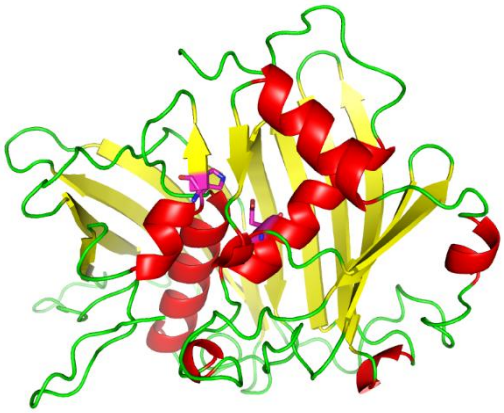


Figure 77 - HHPred model of the SQTKS DH domain

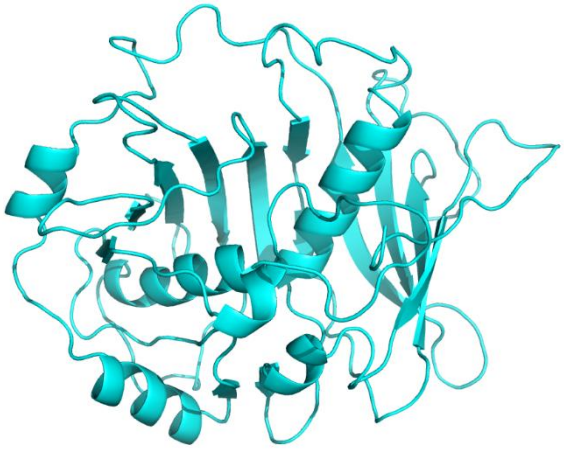


Figure 78 – Swiss Model of the SQTKS DH domain

5.2.3.6 SWISS-MODEL[153]

SWISS-MODEL is a fully automated server with a user friendly web interface which allows the user to generate a protein structure homology model without the use of complex software packages.[153] The SQTKS DH protein sequence was searched against the SWISS-MODEL database to identify possible model templates.[153] The CurFDH domain was returned as the top template match, with the highest sequence identity and coverage. A further model (Figure 78) was subsequently built using SWISS-MODEL's automated modelling server using CurFDH as a template. SWISS-MODEL evaluates the accuracy of a model using a range of parameters. The global model quality estimation (GMQE) score quantifies how accurate a model is and assigns a value between 0-1, the closer to one the more accurate the model. The model was given a GMQE score of 0.53 calculated by SWISS-MODEL.

5.2.4 Comparison of SQTKS DH models

Once all the models had been produced, they were compared in order to identify the most suitable model. Initially the models were overlaid and examined in PyMol. The core structure of the active site in all 6 models was highly similar. The α-helix and β-sheets which make up the double hotdog fold in the DH domain closely overlap in all six models (Figure 79, A-C). In addition the active site amino acids, histidine and aspartic acid, also overlay well in all six models (Figure 79, A). While the active site of all six models are highly similar, the structure of α-helices which are not directly involved in the active site vary between models. In some models the length of the α-helix differs slightly (Robetta model, Figure 79, E), while in others it is the orientation of the α-helix which is a little different. (Phyre2 model, (Figure 79, D).

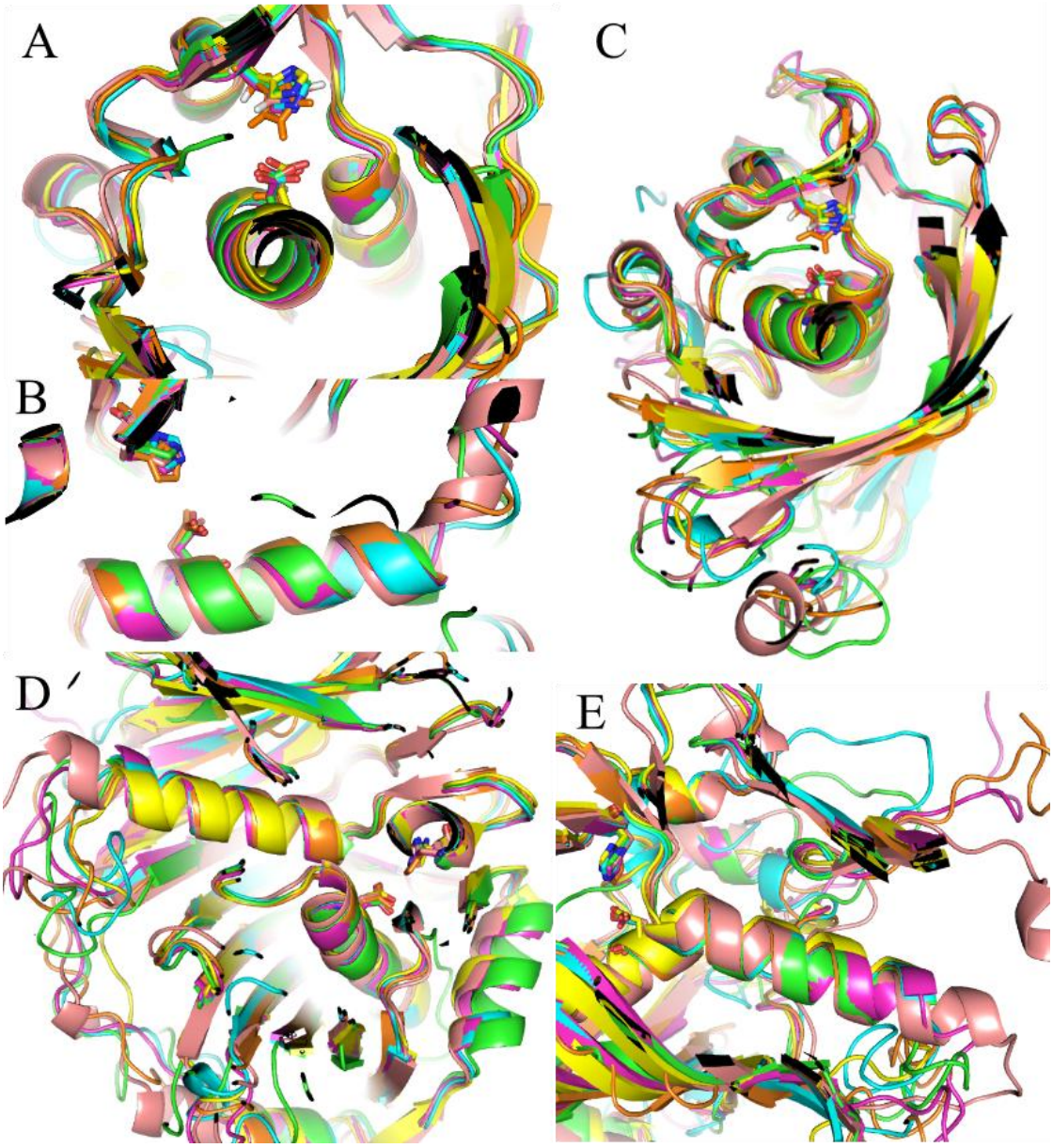


Figure 79 – Overlay for all six models: Swiss Model, green; Itasser model, orange; Phyre, yellow; Raptor X, magenta; Robetta, pink; HHPred, blue; **A**-**B**, Overlay of protein models showing active site α-helix and active site amino acids; **C**, Overlay of protein models showing β-sheets from different model map onto one another; **D**, Overlay of protein models showing variation in α-helix orientation; **E**, Overlay of protein models showing variation in α-helix length.

As all six models were very similar visually, further analysis was undertaken in order to investigate the accuracy of the models and identify the most suitable one. SWISS-MODEL possesses a number of model assessment tools which are designed to assess the accuracy of the model. QMEAN is a combined scoring function which estimates the quality of each residue in the model, was well as the quality of the entire model.[162-164] QMEAN4 evaluates the accuracy of the model by scoring four structural functions including the torsion angle over three amino acids; residue level secondary structure interactions, assessment of the long range interactions; and the solvation potentials.[163] QMEAN6 evaluates the accuracy of the model using the four structure functions used to calculate QMEAN 4 as well as two additional functions;[162] the

agreement prediction of secondary structure (from the sequence) compared with the calculated secondary structure and predicted solvent accessibility.

All the models were submitted the QMEAN server, the QMEAN4 and QMEAN6 scores are summarised in Table 6. A higher QMEAN score indicates a better agreement between the model structure and experimental structures of a similar size. QMEAN scores of -4.0 or below indicate a poor quality model.[164, 165] Itasser was the worst performing of all the models with a QMEAN4 score of -11.25, indicating it is a poor quality model. SWISS-MODEL, RaptorX, HHPred and Phyre2 all scored higher then Itasser but still had QMEAN4 scores below -4.0. The only model to score above -4.0 was the model produced by the Robetta server.

Model	QMEAN4 Score	QMEAN6 Score
Swiss Model	-6.59	-6.01
Itasser	-11.25	-8.73
Phyre2	-6.61	-4.85
RaptorX	-4.66	-3.56
Robetta	-0.75	-0.77
HHPred	-5.55	-4.05

Table 6 - QMEAN4 and QMEAN6 of all six protein models.

In addition to the QMEAN4 and QMEAN6 scores, SWISS-MODEL's evaluation QMEAN server also provides QMEAN Z-scores which allow the models to be compared to high resolution crystal structures in the PDB. The results are reported as comparison plots showing how the protein SQTKS DH model compares to known crystal structures (Figure 80). The comparison plots for all but the Robetta model show the models are dissimilar to high resolution crystal structures in the PDB.

The Robetta model performed the best in the QMEAN testing, the model used the MASDH as a template. The MASDH has the highest sequence homology of all the crystallised DH domains currently available. The Robetta model was therefore taken forward as the model suitable SQTKS DH model in the absence of a crystal structure.

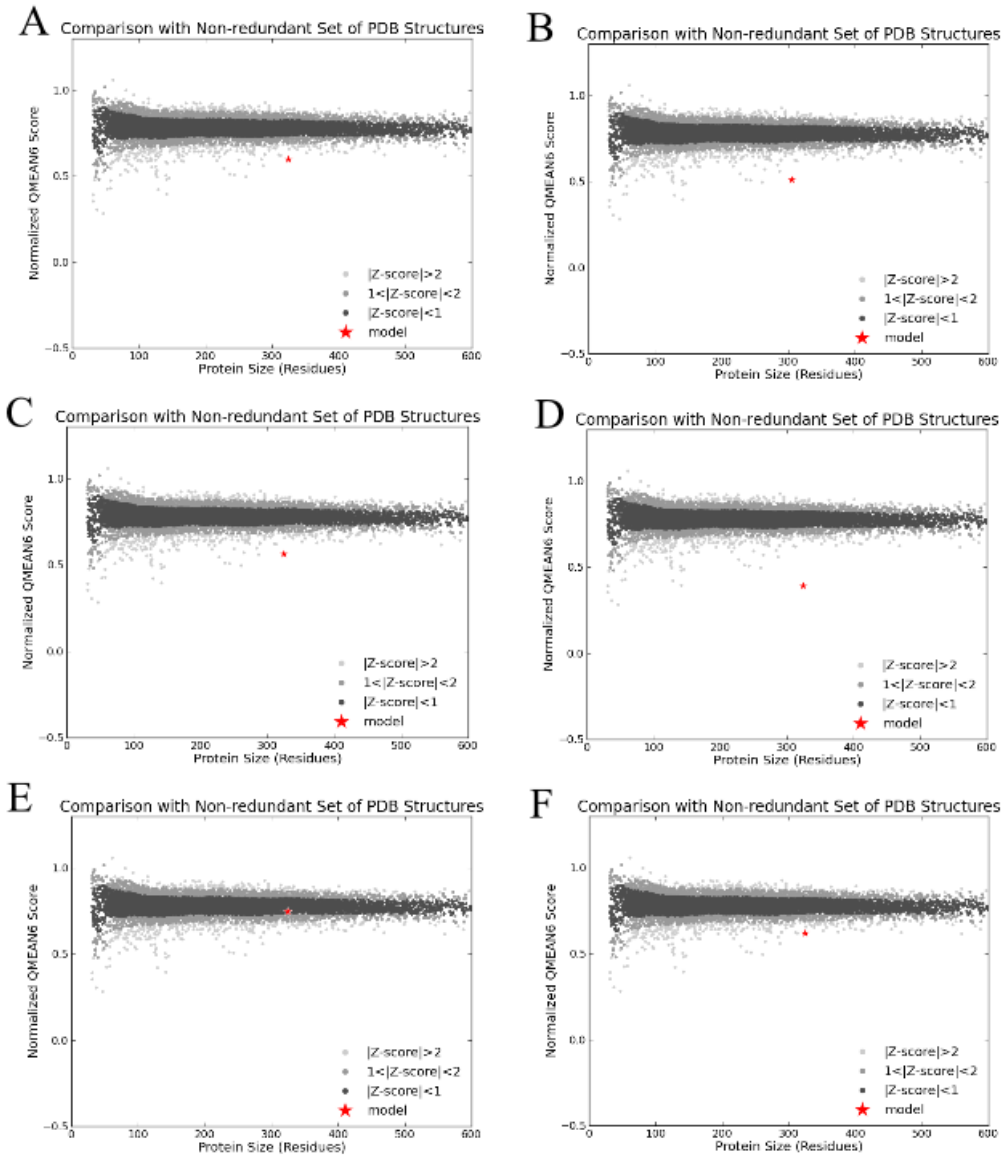


Figure 80 – Comparison plots of protein models, showing QMEAN scores expressed as Z-Scores. Each dot represents a protein structure, the darkest dots are structures with QMEAN Z-scores between 1 and -1. grey dots are structures with QMEAN Z-scores between 1 and 2. Light grey dots represent structures with QMEAN scores greater than 2. The red star represents the model. **A**, HHpred model; **B**, Swiss Model; **C**, Phyre2 model; **D**,Itasser model; **E**, Robetta model; **F**, RaptorX model.

5.3 Docking studies

5.3.1 Diketide Substrate docking

There are four possible natural diketide substrates for the SQTKS, **114**, **115**, **116** and **117** (Figure 81). Investigations into the selectivity of the SQTKS DH domain (Chapter 4) revealed the DH domain was highly selective, with only 2*R*, 3*R*-**114** substrate undergoing dehydration to form **112**. Inhibition studies which were carried out (0) found that the substrate **115**, **116**, and **117**, which had not turned over, were unable to enter the active site of the DH domain. Therefore, only the docking of **114** was studied.

The substrate **114** was initially drawn in ChemDraw[166] and subsequently imported into ChemDraw 3D pro,[166] which allowed the structures to be transformed from 2D image to a 3D representation of the molecules and be saved as a 3D coordinate (PBD) file. Once the substrate was converted into a 3D model, it was imported into PyMol and manually docked into the active site of the Robetta SQTKS DH model. The substrate was roughly aligned with the catalytic amino acids H44 and D235. Once this had been done the PDB files were submitted for energy minimization using YASARA.[167] YASARA force field server carries out molecular dynamic simulations in order to minimize the energy functions of the docked substrate and the protein model, therefore increasing the accuracy of the model.

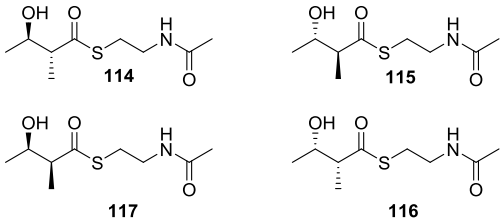


Figure 81 – The four possible diketide substrate mimics.

After undergoing energy minimisation the position of the substrate in the active site was examined (Figure 82). The orientation of the substrate in relation to the active site amino acids was examined and the distances between the α-hydrogen and the catalytic histidine's nitrogen and the β-hydroxyl group and the catalytic aspartic acid were measured (Figure 84).

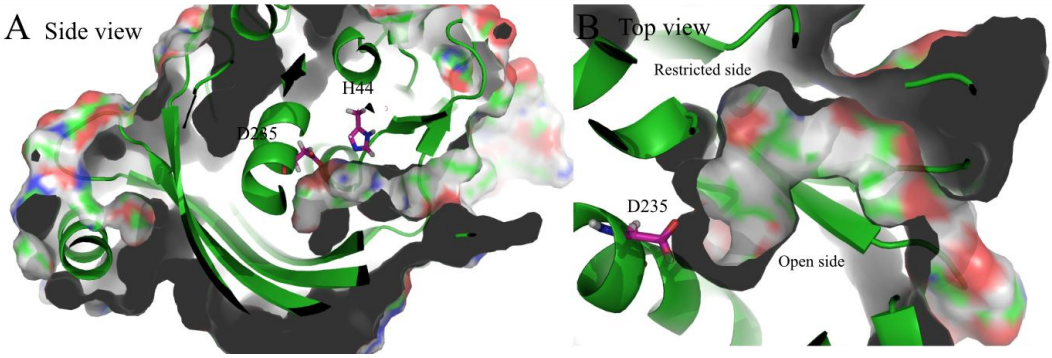


Figure 82 – **A**. Substrate tunnel from the side; **B**. Substrate tunnel viewed from the top

The substrate tunnel is formed by the large central α-helices within the protein, the tunnel runs along the length of the C-terminal α-helix and curves towards end where the active site is situated. One side of the substrate tunnel, the side nearest the catalytic aspartic acid, is restricted making it unable to accommodate large bulky groups (Figure 82). The opposite side is much more open and widens around the active site, so it can accommodate the growing polyketide chain.

The substrate 2*R*,3*R*-**114** is shown docked into the active site of the DH in Figure 84 A and B. 2*R*,3*R*-**114** comfortably fits into the active site, with the α-methyl and β-methyl groups occupying a large open pocket within the active site. The positioning of the methyl groups in this way allows the α-hydrogen to be close enough for the catalytic histidine's nitrogen to abstract the α-hydrogen, therefore facilitating the elimination reaction (3.5 Å, Table 7). The β-hydroxyl group is at a similar distance, where a proton from the catalytic aspartic acids hydroxyl group can hydrogen bond to the oxygen of the β-hydroxyl group (2.6 Å, Table 7).

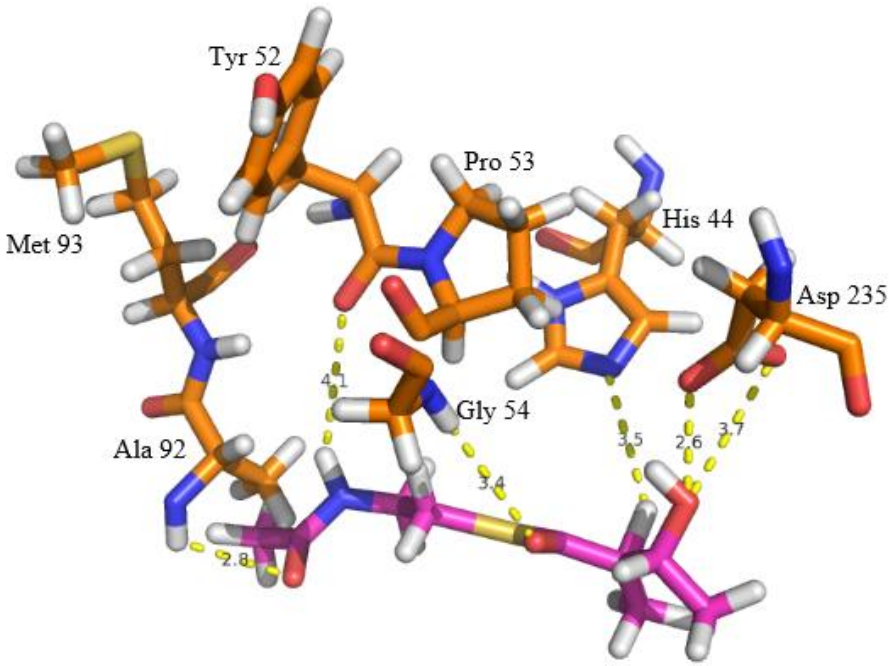


Figure 83 – Key modelled interactions between the substrate 3*R*, 3*R*-**114** (magenta) and key amino acids with in the SQTKS DH active site (orange).

The docked 2*R*, 3*R*-**114** substrate is able to interact with other amino acids with in the active site. The thioester carbonyl is able to hydrogen bond to the NH of a nearby glycine (G54, 3.4 Å, Figure **83**), helping to stabilise the positive charge which will form on the carbonyl during the course of the reaction. The SNAC NH can hydrogen bond to the backbone carbonyl of a nearby tyrosine (Y52, 4.1 Å). The SNAC carbonyl is situated close to a nearby alanine (A92, 2.8 Å). The interactions of these amino acids help position the SNAC correctly in the active site. It can be presumed that the amino acids would interact in a similar fashion with a pantetheinyl-ACP bound substrate. The positioning of the α-hydrogen and β-hydroxyl group within hydrogen

bonding distance of the catalytic residues indicates that the substrate will be able to undergo dehydration in the presence of the SQTKS DH enzyme. This finding is supported experimentally, the assay of 2*R*,3*R*-**114** underwent dehydration when in the presence of the DH enzyme.

Substrate	α/Å	β/Å
2*R*,3*R*-114	3.5	2.6
3*R*-124	2.6	2.6
3*R* intermediate	-	1.6

Table 7 - Measured distances between α-proton and histidine nitrogen and between β-hydroxyl group and aspartic acid hydroxyl proton

The active non-methylated substrate **124** was also docked into the active site (Figure 84). The substrate **124** fitted well within the active site and was able to interact with catalytic amino acids within hydrogen bonding distance, (α-hydrogen, 2.6 Å, β-hydroxyl, 2.6 Å, Table **7**).

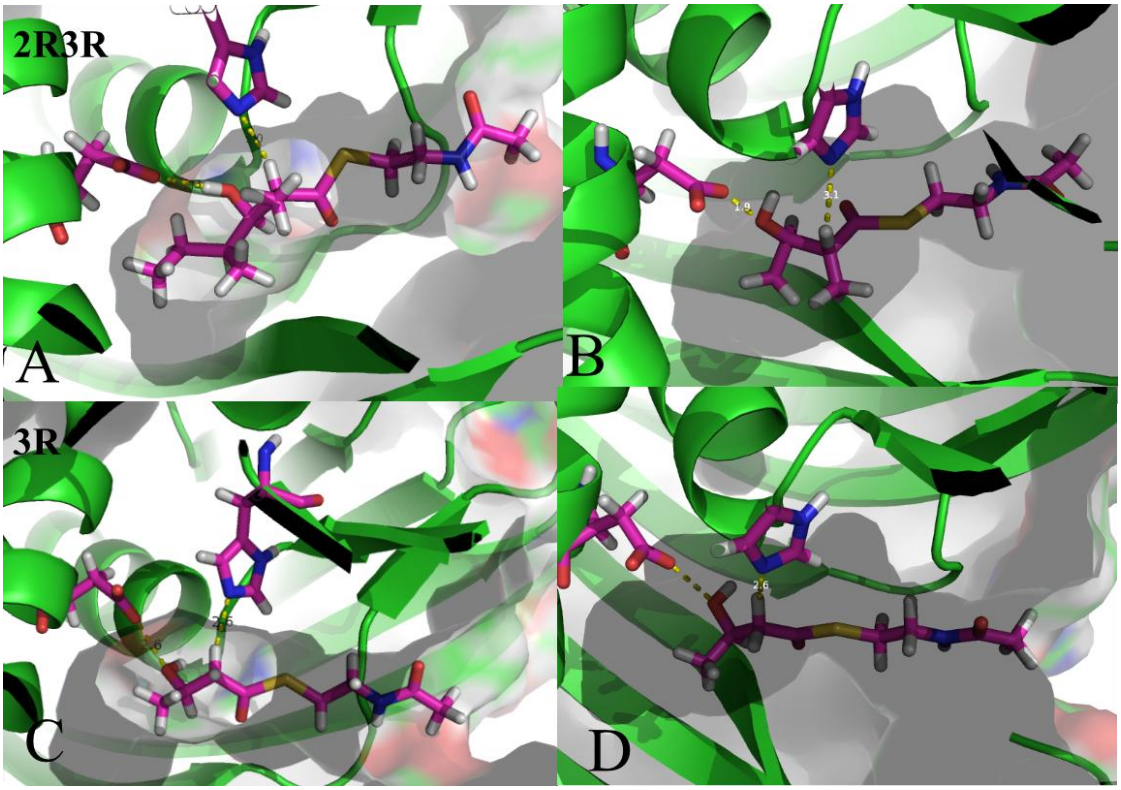
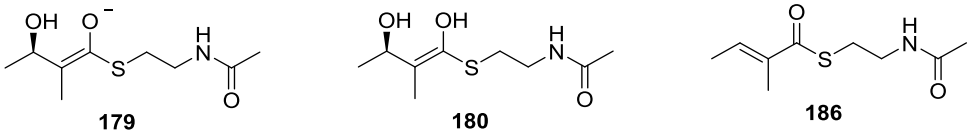


Figure 84 – Docking of the active substrate mimics within the active site of the Robetta SQTKS model **A-B**, docking of the methylated substrate 2*R*,3*R*-**114**; **A**, side view, **B**, top view; nonmethylated substrate mimic with in the active site of the CurF-SQTKS model. **C-D**, docking of the nonmethylated 3*R*-**124**; **C**, side view, **D**, top view;

5.3.2 Diketide intermediate docking

To complement the diketide substrate docking experiments, the diketide enolate **179** and enol **180** intermediates were also docked into the active site. Both the enol and enolate were placed into the active site using PyMol and roughly aligned with the catalytic amino acids H44 and D235. These structures were then subsequently submitted for energy minimization using YASARA.[167] During the energy minimization process YASARA transformed the enolate into the enol, as it was deemed the most stable form, it was not possible to prevent this transformation, and therefore only the enol will be discussed further.



The positioning of the enol diketide substrate **180** within the active site is comparable to the positioning of the 2*R*,3*R*-**114**. The distance between the β-hydroxyl and D235 is reduced from 2.7Å to 1.6Å in this docking simulation (Figure 85). The proposed mechanism is discussed in chapter 1, Section **1.5**, Scheme 13. The α-methyl group is now in a planar position, allowing the methyl group to comfortably fill the open side of the active site tunnel. The planar positioning of the methyl group appears to allow the β-hydroxyl group to be positioned closer to D235, it is possible that this helps facilitate the loss of water needed to complete the dehydration reaction. Once the dehydration has occurred the substrate is released from the active site, as it no longer possesses a hydroxyl group and therefore is unable to hydrogen bond to D235.

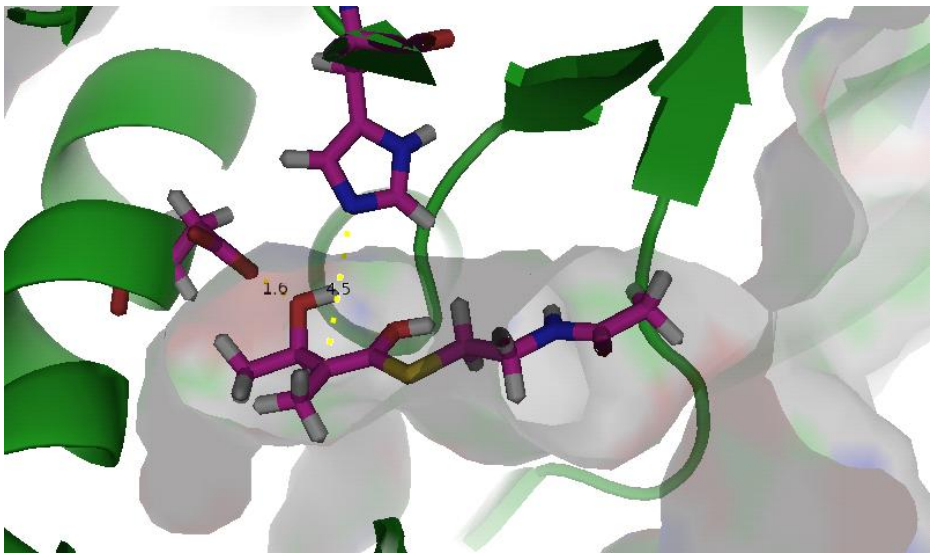
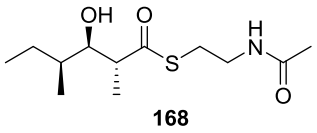


Figure 85 – Docking of diketide enol intermediate

5.3.3 Triketide docking

The triketide substrate mimic **168**, was hypothesised to be an active substrate and was docked using the same process used to dock diketide substrates in Section 5.3.1. The substrate was docked into the active site of the Robetta SQTKS model and the α-proton and β-hydroxyl groups were roughly aligned with H44 and D235. The docking model was then subjected to energy minimization, carried out by the YASARA server. [167] The conformation of the substrates within the active site was then examined and the distance from the catalytic amino acids, H44 and D235 to the α-proton and β-hydroxyl group was noted.



168

Figure 86 – Triketide substrate mimic.

As with the diketide substrate mimics, the stereochemistry of the substrate dictates how the substrate fills the active site and interacts with the catalytic amino acids. The 2*R*,3*R*,4*S* substrate **168** (Figure 87, A) docks into the active site with the α-methyl group positioned away from the restricted side of the substrate tunnel, towards the more open face of the tunnel. This allows the α-proton and β-hydroxyl group to locate within hydrogen bonding distance. The α-proton is 2.7 Å away from H44 and β-hydroxyl is 2.7Å away from D235. The long alkyl chain is threaded through the substrate tunnel and curves down into the open pocket at the end of the tunnel.

The α-proton and β-hydroxyl and catalytic amino acids interact at a distance which indicates that abstraction of the α-proton by H44 and protonation of the β-hydroxyl group is possible. The docking of the 2*R*3*R*4*S*-**168** indicates that it is likely be an active substrate and will undergo dehydration.

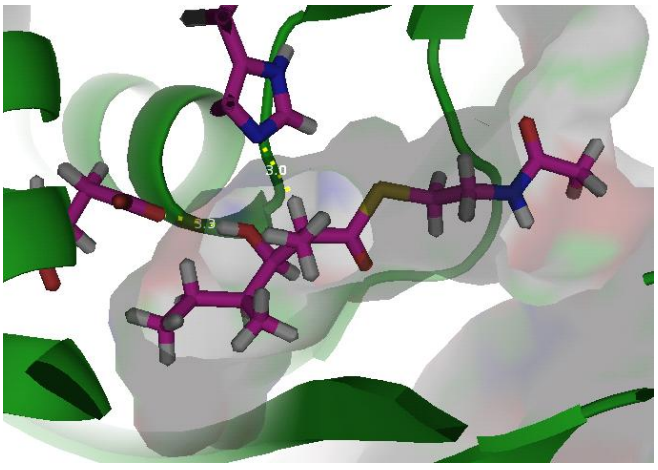


Figure 87 – 2*R*,3*R*,4*S*-**168** triketide substrate mimic docked into the active site of Robetta SQTKS model

5.4 Conclusion

In the absence of an available crystal structure six SQTKS DH protein models were produced using a range of modelling programs which were tested using SWISS-MODEL QMEAN server.

The model produced by the Robetta program was deemed the most suitable and used MASDH as a template with which to build the model. Although the sequence identity between MASDH and SQTKS was lower than desired (25.5%), of all the crystallised DH domains, MASDH has the highest sequence identity. The Robetta model was also the only model to have an QMEAN score above -4.

Studies of crystal structures of other DH domains indicate that the difference in substrate size should not affect the modelling outcomes to any great degree. The substrate tunnels of other DH domains appear to lack amino acids which are able to carry out hydrogen bonding for explicit substrate recognition, as the substrate is always delivered by the ACP.[6] The structural similarities observed in these structures indicate that SQTKS DH will also possess a substrate tunnel lacking in hydrogen bonding amino acids. The active site contains two amino acids which are able to hydrogen bond to the substrate in such a way as to facilitate the dehydration reaction, which are a histidine and an aspartic acid residue. Additional amino acids are able to hydrogen bonding to the substrate, it is though that these amino acids help stabilise the substrate with in the active site during the dehydration reaction. A number of other amino acids in the active site is involved in hydrogen bonding to the catalytic amino acids in order to stabilise and increase their activity.

The docking studies indicate that the programming in the DH domain arises from the positioning of the catalytic amino acids and the shape of the substrate tunnel. The mechanism of dehydration is reliant on the abstraction of proton from the α-position and protonation of the β-hydroxyl group, and subsequent loss of water. This results in strict requirements for the dehydration to occur. The substrate must be able to interact with the catalytic amino acids within hydrogen bonding distance. The shape of the tunnel and the positioning of the amino acids dictates which stereoisomer substrate is tolerated and turned over.

Incorrect stereochemistry results in the substrate being unable to interact with the catalytic amino acids at the required distance and therefore unable to undergo dehydration.

Chapter 6 - Conclusion and future work

The focus of the work described in this project was to investigate the selectivity and kinetic characteristic of the SQTKS DH domain.

In order to probe the programming of the SQTKS active site a range of diastereometic and enantiomerically pure substrates were synthesised. Four natural diketide substrates **115** - **117** were synthesised, the *syn* diastereomers were synthesised utilising Evans auxiliaries to control the stereocenters formed during the aldol reaction. The *anti* diastereomers stereocenters were installed using Fráter-Seebach alkylation. In addition to the natural diketide substrates two non-natural substrates were also synthesised, they were designed to assess the importance of the β-methyl group to substrate turn over.

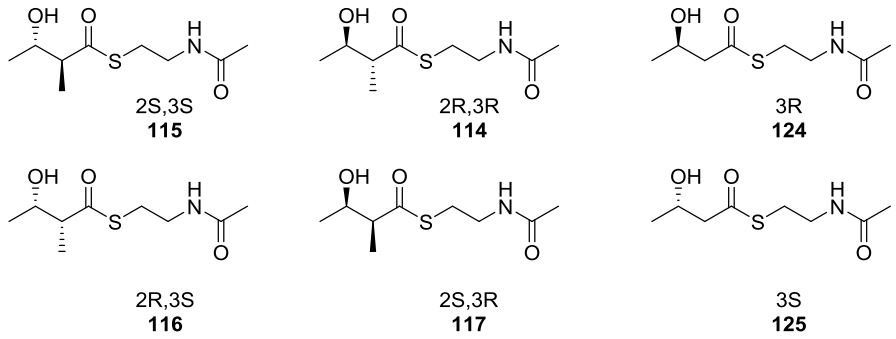
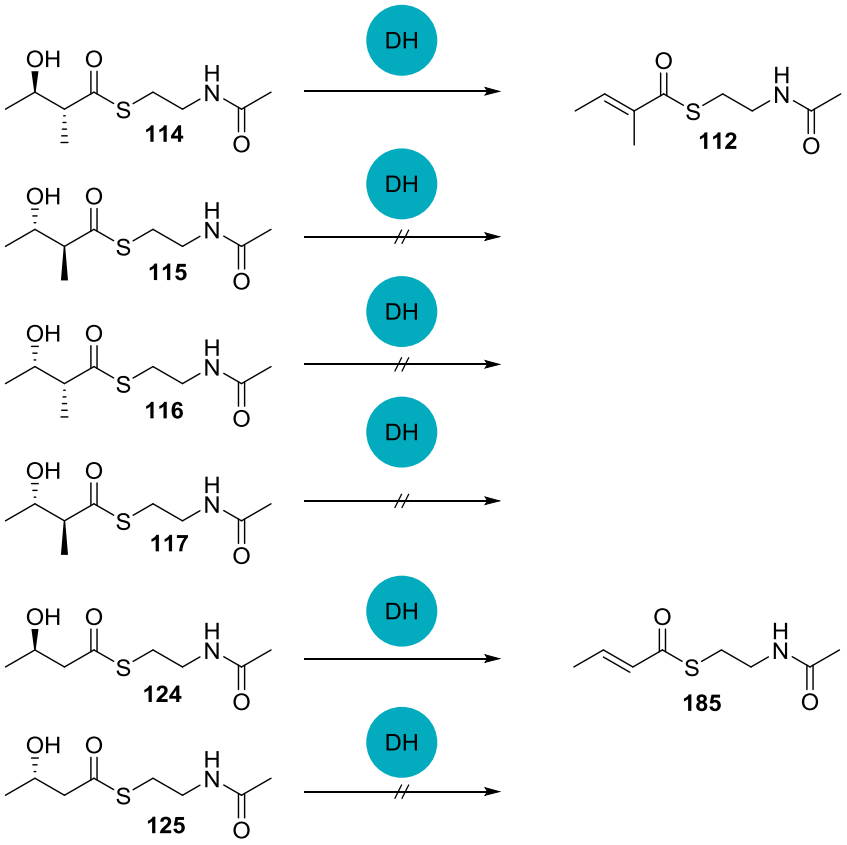


Figure 88 – synthesis diketide mimics.

While it was possible to produce the SQTKS DH as soluble protein, the enzyme was found to be unstable. The optimisation of purification and storage buffers allowed the DH domain to be stabilised in solution by L-Arginine and L-Glutamic acids. Despite additives to help stabilise the protein, precipitation still occurred over time, but it was possible to store the enzyme at 4 °C for up to 2 months and carry out the required assays.

The assays were designed to test the selectivity and kinetic characteristics of the isolated DH towards diketide substrates mimics. The assays progress was monitored by measuring the formation of the tigloyl-SNAC product *via* SIM LCMS. The assays demonstrated the DH domain was highly selective, only catalysing the dehydration of one natural substrate mimic **114** out of the possible four stereoisomers. The DH domain was also able to act on one non-natural substrate **124**. The investigation found that substrate 2*R*,3*R*-**114** was processed more efficiently than by the isolated DH 3*R*-**124**.

Inhibition studies showed no significant reduction in the rate **114** is processed by the DH when incubated in the presence of an additional substrate (**115**, **116** and **117**). These results indicated that is it highly likely the **115**, **116** and **117** are unable to enter the DH active site.



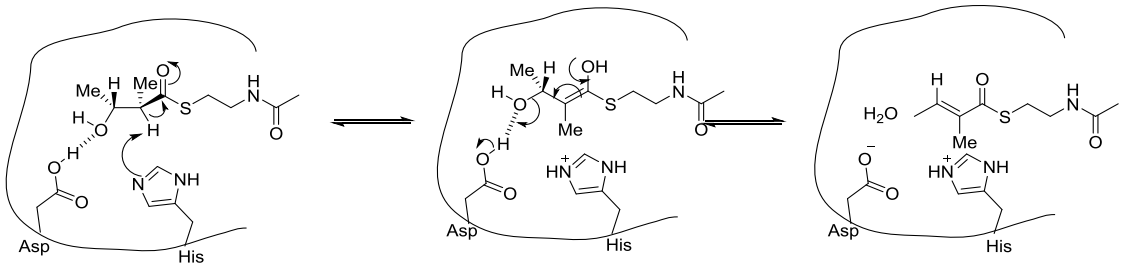
Scheme 59– Selective exhibited by the SQTKS DH domain towards methylated and non-methylated substrate mimics.

Attempts at crystallizing the DH domain were unsuccessful. In the absence of a crystal structure, a model was created to gain insight into how the active substrate sit within the active site. PKS DH domains share limited sequence identity but when the crystal structures are compared, they share a high degree of structural similarity. To limit the inaccuracies which could arise due to the low sequence identity between the SQTKS DH domain and the available PKS crystal structures a range of models were created. The models were assessed using SWISS-MODEL QMEAN software and the most suitable model was produced by the Robetta server. This model DH structure was then used in docking studies with various possible substrates. Only the active substrates were docked into the active site of the model as the inhibition studies had shown the non-substrate did not enter the active site.

The results from the docking study fit well with the results observed in the chemical assay, rationalising the observed stereoselectivity of the DH domain. The assays showed the DH domain programming controls which stereoisomer is accepted and subsequently dehydrated (Scheme 59). The mechanism for DH requires the substrate to interact with two catalytic amino acids, histidine and aspartic acid. The model of SQTKS DH rationalised the hypothesis that the programming of the DH arises from the positioning of the amino acids within the active site (Scheme 60). The positioning in turn dictates which stereoisomer is tolerated. For a substrate to be dehydrated it must contain a 3*R*-hydroxyl group. The DH domain can act on non-methylated

substrate, but if a methyl group is present in the α-position then it must also have *R* stereochemistry.

The modelling also showed additional amino acids within the active site tunnel can interact with key sections of the substrate. These amino acids help position and stabilise the substrate, so the dehydration reaction can occur. The nearby glycine which interacts with thioester carbonyl, can stabilise negative charges which form temporarily during the reaction.



Scheme 60 – Dehydration mechanism of the 2R,3R-**114** substrate with in the DH active site.

The SQTKS DH domains selectivity also fits well with the selectivity observed in other known DH enzymes. The crystallised DH domains, Ery4DH and Rif10DH both exhibit the same selectivity towards the 2*R*,3*R*-2-methyl-3-hydroxyacyl-SNAC substrates.[12, 13] The selectivity towards the 3*R*-hydroxyl group is also observed in mFAS,[168] yeast FAS,[24] bacterial FabA and FabZ.[70, 71] All of the aforementioned DH domains catalyse the dehydration to give the *Z* geometry α,β-unsaturated thiolester. Docking also showed how the *syn* elimination geometry is controlled which agrees with the observation that the 2*R*,3*R*-**114** substrate gives the *E* product.

Further work on the SQTKS DH will focus on the interactions between the DH and the ER domain. The investigations will examine how interactions between the DH and ER domain affect the selectivity's and kinetic characteristic of both domains. Also, a larger fragment of the SQTKS, from the DH to KR has successfully been cloned and isolated. Work will be carried out on the larger fragment to assess its selectivity.

The interaction between the "non-substrates" and the DH and ER domain will also be studied using surface plasmon resonance (SPR). The aim of the investigation is to find out if the alcohol substrates can enter the active site of the ER and if the alkene substrates are able to enter the DH active site. If the substrates do enter the active site, the use of SPR will also allow the time substrates spend in the active site to be monitored.

Chapter 7 – Experimental

7.1 General Microbiological techniques

7.1.1 Media

All growth media were sterilised by autoclaving at 121 °C for 10 minutes prior to use.

For solid medium, *E. coli* was grown on LB-agar containing 10 g/L tryptone, 5 g/L NaCl and 15 g/L agar. Sterile agar was melted in a microwave and allowed to cool slightly. Appropriate antibiotics were added, and the molten agar poured into sterile petri dishes either in a laminar flow hood or under sterile conditions.

SOC recovery medium used during *E. coli* transformation, contained 0.5 % (w/v) yeast extract, 2 % (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

LB Glu/Carb agar containing 10 g/L tryptone, 5 g/L NaCl and 15 g/L agar. Sterile agar was melted in a microwave and allowed to cool slightly. 2 mL of 20 % glucose stock and 200 µL of 100 mg/mL carbenicillin stock per 200 mL were added and the molten agar poured into sterile petri dishes either in a laminar flow hood or under sterile conditions.

LB liquid medium was used for starter cultures of *E. coli*. It contained 10 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl.

2TY medium used for growth of *E. coli* for protein expression contained 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl.

2TY-glucose medium used for growth of *E. coli* for protein expression contained 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl. After sterilisation,5 % glucose w/v was added.

2TY Enhanced (2TYE) medium for growth of *E. coli* used for protein expression contained 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl dissolved in 900 mL of H_2O. After sterilisation 1 mL/L 2 M $MgSO_4$, 1 mL/L metal mix A, 100 mL/L 100 mM phosphate buffer and 0.5 % glucose w/v were added.

Terrific Broth (TB) medium used for growth of *E. coli* for protein expression contained 12 g/L Bacto-tryptone, 24 g/L Bacto-yeast extract, 4 mL/L glycerol dissolved in 900 mL of H_2O. After sterilisation 100 mL 100 mM phosphate buffer was added.

Terrific Broth Enhanced (TBE) used for growth of *E. coli* for protein expression contained 12 g/L Bacto-tryptone, 24 g/L Bacto-yeast extract, 4 mL/L glycerol dissolved in 900 mL of H_2O. After sterilisation 100 mL/L 100 mM phosphate buffer, 1 mL/L metal mix A and 0.5 % glucose w/v were added.

Rich media was used for growth of *E. coli* for protein expression. It contained 2TY medium, which contained 16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl. After autoclave sterilisation the following autoclave sterilised components were added; 50 ml/L 20xNPS, 1 mL/L metal mix B. The following filter sterilised components were also added; 25 ml/L glucose, glycerol, $MgSO_4$ stock solution.

Auto induction medium used for growth of *E. coli* for protein expression contained phosphate buffer (pH 7.2) consisting of 6 g/L Na_2HPO_4 and 3 g/L KH_2PO_4. 20 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl. After autoclave sterilisation the following filter sterilised components were added; 10 mL/L 60 % v/v glycerol, 5 mL/L 10 % w/v glucose and 25 mL/L 8% w/v lactose.

Auto induction-LB5052 medium used for growth of *E. coli* for protein expression contained 5 g/L yeast, 10 g/L tryptone, 5 g/L glycerol, 0.7 g/L Na_2SO_4, 2.5 g/L NH_4Cl dissolved in 900 mL H_2O. After sterilisation 0.5 g/L glucose, 2 g/L lactose, 1 mL/L 2 M $MgSO_4$, 1 mL/L metal mix A, 100 mL/L 50 mM phosphate buffer and 0.5 % glucose w/v were added.

20xNPS contained 6.6 g/L NH_4SO_4, 13.6 g/L KH_2PO_4, 14.2 g/L Na_2HPO_4. glucose, glycerol, $MgSO_4$ stock solution contained 5 g/100 mL glucose, 25 g/100 mL glycerol and 8 mL 1M $MgSO_4$

Metal mix A stock solution was made as follows: solutions of the metal compounds, tabled below were made in water at their defined volume and concentrations. $FeCl_3.6H_2O$ was made up in 50 mL 0.1 M HCl instead of distilled water. All metals dissolved in distilled water were added to 46 mL of distilled water and filter sterilised. The $FeCl_3.6H_2O$ in 50 mL 0.1 M HCl was then added to the filter sterilised solution.

Reagent	**Mass (g)**	**Volume (mL)**	**Concentration (M)**
$FeCl_3.6H_2O$	1.35	50 (0.1 M HCl)	0.1
$MnCl.4H_2O$	0.91	1	1
$ZnSO_4.7H_2O$	0.28	1	1
$CoCl.2H_2O$	0.23	1	0.1
$NiCl_2.6H_2O$	0.23	1	0.2

Metal Mix B stock solution was made as follows: solutions of the metal compounds, tabled below were made in water at their defined volume and concentration. $FeCl_3.6H_2O$ was made up in 50 mL 0.1 M HCl instead of distilled water. The solutions were added to distilled water (36 mL) and autoclave sterilised.

Reagent	Mass (g)	Volume (mL)	Concentration (M)
$FeCl_3.6H_2O$	1.35	50 (0.1 M HCl)	0.1
$CaCl_2$	0.22	2	1
$MnCl.4H_2O$	0.20	1	1
$ZnSO_4.7H_2O$	0.29	1	1
$CoCl.2H_2O$	0.05	1	0.2
$CuCl.2H_2O$	0.03	2	0.1
$NiCl_2.6H_2O$	0.05	1	0.2
$NaMoO_4.2H_2O$	0.05	2	0.1
$NaSeO_3.5H_2O$	0.05	2	0.1
H_3BO_3	0.012	2	0.1

Phosphate buffers were made as follows using distilled water and filter sterilised before use. 100 mM phosphate buffer consisted of 200 mL/L 1M KH_2PO_4 and 800 mL/L 1M K_2HPO_4. 50 mM phosphate buffer consisted of 100 mL/L 1M KH_2PO_4 and 400 mL/L 1M K_2HPO_4 and 500 mL/L distilled water.

7.1.2 Stock solutions

7.1.2.1 Antibiotic stock

0.2 M carbenicillin stock contained 1 g/ 10 mL of carbenicillin in H_2O, which was filter sterilised using a 0.2 µL filter, divided into aliquot and stored at -20 °C.

7.1.2.2 Induction stock solutions

20 % (w/v) rhamnose solution contained 2 g/ 10 mL of rhamnose in H_2O, which was filter sterilised using a 0.2 µL filter, aliquot and stored at room temperature.

1 M IPTG stock contained 2.3 / 10 mL of IPTG in in H_2O, which was filter sterilised using a 0.2 µL filter, aliquot and stored at -20 °C.

20 % (w/v) glucose stock contained 2 g/ 10 mL of glucose in H_2O, which was filter sterilised using a 0.2 µL filter, aliquot and stored at room temperature.

7.2 Transformation protocol

Frozen 25 µL aliquots of the appropriate *E. coli* strain were defrosted on ice. 0.5 µL of plasmid DNA was added and incubated on ice for 30 minutes. The cells were heat-shocked at 42 °C for 30 seconds and immediately placed onto ice. After resting on ice for 2 minutes, 200 µL SOC medium was added and the tubes were incubated in a water bath at 37 °C for 1 hour. 100 µL portions of the transformation mixture were spread into LB agar plate containing the appropriate antibiotic(s) and the plates incubated at 37 °C. Agar plates for transformation of *E. coli* BL21

(DE3) contain (1 mL/L) carbenicillin. Agar plates for transformations of *E. coli* Rosetta strain contain (1 mL/L) carbenicillin and (1 mL/L) chloramphenicol.

7.3 Standard expression protocol

Following transformation with the appropriate expression plasmid, a single transformed *E. coli* colony was used to inoculate a 50 mL sterile LB starter culture medium containing appropriate antibiotic(s). This was grown to stationary phase overnight at 37 °C with shaking. This starter culture was used to inoculate auto induction medium supplemented with the appropriate antibiotic(s) at 1:100 dilution in 500mL flasks each containing 200 mL medium. The flasks were incubated at 37 °C with shaking until the OD_{600} reached 0.6.

The temperature of the culture was then adjusted to the expression temperature of 19 °C and induced with the appropriate inducer (1 M IPTG or rhamnose) solution which was added with the appropriate final concentration. The cultures were incubated with shaking overnight and cells were harvested by centrifugation. The cell pellets were suspended in binding buffer.

7.4 Optimised expression protocol

E. coli BL21 gold (DE3) was used as a host for protein expression. Following transformation with the appropriate expression plasmid using LB-Glu/Car agar plates, a single colony was used to inoculate a 50 mL sterile LB starter culture medium containing appropriate antibiotic(s). This was grown to stationary phase overnight at 30 °C with shaking. This starter culture was used to inoculate auto induction medium supplemented with the appropriate antibiotic(s) at 1:100 dilution in 500mL flasks each containing 200 mL of LB medium. The flasks were incubated at 37 °C with shaking for 4 hours. The temperature of the culture was then adjusted to the expression temperature of 16 °C. Induction when needed was induced with 0.4 mM IPTG. The cultures were incubated with shaking for a further 4 hours and cells were harvested by centrifugation. The cell pellets suspended in binding buffer containing protease inhibitor cocktail and were refrigerated overnight before being purified the following day.

7.5 Protein purification

7.5.1 SDS-Page analysis of protein

Proteins were analysed on 12.5 % tris-glycine polyacrylamide gels using a Bio-Rad Mini-Protean^{TM}gel casting and electrophoresis system. A separating gel was poured between glass plates, overlaid with water-saturated butanol to create a flat surface. Once the gel had set, the butanol was removed, and the stacking gel solution was poured to the top of the gel and a 15 well comb was inserted. Gel solutions were prepared immediately before use as follows:

Component	Volume (mL)	Final concentration
Stacking gel		
1 M tris-HCl pH 6.8	0.25	0.125 M
10 % SDS	0.02	0.1 %
40 % acrylamide/bisacrylamide 37.5:1	0.25	5 %
H_2O	1.46	-
10 % ammonium persulphate	0.02	0.1 %
TEMED	0.002	6.6 mM
Separating Gel		
1.5 M tris-HCl pH 8.8	1	0.375 M
10 % SDS	0.04	0.1 %
40 % acrylamide/bisacrylamide 37.5:1	1.25	12.5 %
H_2O	1.67	-
10 % ammonium persulphate	0.04	0.1 %
TEMED	0.002	13 mM

Gels were placed in the electrophoresis tank and SDS-PAGE running buffer was poured around the gels. For Tris-glycine gels, a single running buffer was prepared from a 10x concentrate. The running buffer contained tris base 68 g/L, glycine 144 g/L, SDS 10 g/L.

Samples from electrophoresis were mixed with 5x SDS-PAGE loading buffer and heated at 95 °C for 5 minutes.

Component	Concentration
Tris-HCl pH 6.8	50 mM
Glycerol	40 %
SDS	1 %
β-mercaptoethanol	10 mM
Bromophenol blue	%

7.5.2 Purification buffers

Binding buffer contained 20 mM Tris-HCl pH8, 150 mM NaCl, 5 mM imidazole, 10 % glycerol, 100 mM L-Arginine and L-Glutamic acid.

Elution buffer contained 20 mM Tris-HCl pH8, 150 mM NaCl, 400 mM imidazole, 10 % glycerol 100 mM L-Arginine and L-Glutamic acid.

Desalting buffer contained 20 mM Tris-HCl pH8, 150 mM NaCl, 10 % glycerol 100 mM L-Arginine and L-Glutamic acid.

7.5.3 Ni^{2+} affinity purification protocol

The cell pellet from *E. coli* cultures was suspended in elution buffer (35-50 mL for each L of culture). The cell suspension was sonicated in a glass beaker on ice at full power for 30 second bursts with 30 seconds rest between bursts for a total of 6 minutes. The lysate was clarified by centrifugation (17,000 RPM for 15 minutes).

A 25 mL his-trap (GE healthcare) column was attached to FPLC instrument and charged with 100 mM $NiSO_4$ and equilibrated with 200 mL binding buffer. The clarified lysate was then loaded onto the column using an FPLC super loop. The column was then washed with enough of binding buffer for the UV trace to settle. The bound protein was eluted over a gradient of 0-100 % elution buffer in 100 mL. 5 mL fractions were collected. Fractions for which a peak in the UV_{280} absorbance trace was observed were analysed by SDS-PAGE to determine the purity and size of the eluted protein.

Binding buffer contained 20 mM Tris-HCl pH8, 150 mM NaCl, 5 mM imidazole and 10 % glycerol. Elution buffer contained 20 mM Tris-HCl pH8, 150 mM NaCl, 400 mM imidazole and 10 % glycerol.

7.5.4 Desalt purification protocol

After Ni^{2+} affinity purification the protein was concentrated using a Vivaspin 20 centrifugal concentrator with a 10 kDa cutoff. A HiPrep^{TM} 26/10 desalting column was equilibrated with 1.5 column volumes of desalting buffer (20 mM Tris-HCl pH8, 150 mM NaCl, 10 % glycerol 100 mM L-Arginine and L-Glutamic acid). Up to 10 mL of protein (for 100 mL HiPrep^{TM} 26/10 column) was injected and eluted over one column volume. When the beginning of a peak in UV_{280} absorbance trace was observed one fraction was collected.

7.5.5 Size exclusion purification protocol

Protein for size exclusion chromatography was concentrated using a Vivaspin 20 centrifugal concentrator with a 10 kDa cutoff. A supersex S300 column was equilibrated with 1.5 column volumes of desalting buffer (20 mM Tris-HCl pH8, 150 mM NaCl, 10 % glycerol 100 mM L-Arginine and L-Glutamic acid). Up to 2 mL of protein (for 120 mL S200 column) was injected and eluted over one column volume. 5 mL fractions were collected and those for which a peak in UV_{280} absorbance trace was observed were analysed.

7.6 DH enzyme assay

7.6.1 DH protein production

E. coli BL21 (DE3) were transformed using the expression plasmid #7615 pOpinF as detailed in section 7.2 and expressed as detailed in section 7.4 DH protein was prepared by Ni^{2+} affinity and size exclusion chromatography as described in section 7.7.1.

7.6.2 DH substrate specificity enzyme assay

DH substrate specificity enzyme assays were performed in a volume of 100 µL and contained the following:

Component	Volume (µL)	concentration
1 M Tris-HCl pH8	10	50 mM
50 % Glycerol	40	33 %
Substrate mimic	10	~10 mM
DH (10 mg.mL)	50	132 µM

The mixture was incubated at 30 °C for 8 hours. For analysis, at set time points, 20 µL aliquots were taken and mixed with 60 µL of acetonitrile, 20 µl H_2O, then centrifuged in a microcentrifuge to remove the precipitated protein. The supernatant was removed analysed by LC-MS as described in section 7.7.1..

7.6.3 DH inhibition assays

DH substrate specificity enzyme assays were performed in a volume of 100 µL and contained the following:

Component	Volume (µL)	concentration
1 M Tris-HCl pH8	10	50 mM
50 % Glycerol	40	33 %
Substrate (2*R*,3*R*)	5	2.00 mM
Substrate (2*S*,3*S*) or (2*R*,3*S*) or (2*S*,3*R*)	5	2.00 mM
DH (5 mg.mL)	50	66 µM

The mixture was incubated at 30 °C for 1 hour. For analysis, at set time points, 20 µL aliquots were taken and mixed with 60 µL of acetonitrile, 20 µl H_2O, then centrifuged in a microcentrifuge to remove the precipitated protein. The supernatant was removed and analysed by LC-MS as described in section 7.7.1.

7.6.4 DH Kinetic enzyme assay

DH substrate specificity enzyme assays were performed in a volume of 100 µL and contained the following:

Component	Volume (µL)	concentration
1 M Tris-HCl pH8	10	50 mM
50 % Glycerol	40	33 %
Substrate mimic	10	0.2-3.99 mM
DH (mg.mL)	50	132 µM

The mixture was incubated at 30 °C for 1 hours. For analysis, at set time points, 20 µL aliquots were taken and mixed with 60 µL of acetonitrile, 20 µl H_2O, then centrifuged in a microcentrifuge to remove the precipitated protein. The supernatant was removed and analysed by LC-MS as described in section 7.7.1.

7.7 General synthetic techniques

7.7.1 Synthetic methods

All reagents were purchased from commercial suppliers and used without further purification, unless otherwise stated. Common anhydrous solvents were obtained by passing through a modified Grubbs system of alumina columns, manufactured by Anhydrous Engineering.[169] DMF was purchased as an anhydrous solvent. Triethylamine and diisopropylamine were distilled over calcium hydride and stored under an inert nitrogen atmosphere.

All moisture or air sensitive reactions were carried out in flame dried glassware under a positive pressure of nitrogen using standard Schlenk-line techniques. After aqueous work of reaction mixtures, organic solutions were routinely dried with anhydrous magnesium sulphate.

Column Chromatography

Column chromatography was performed using Merck Kieselgel 60 silica gel, eluting with the solvent system stated. Columns were eluted under pressure and the fractions analysed by TLC. Fractions with the same R_f were pooled and concentrated *in vacuo*.

NMR

NMR spectra were recorded at 25 °C using Varian 400-MR or Jeol ECP (Eclipse) 400 spectrometers at 400MHz for ^{1}H and ^{13}C spectra as stated. Proton spectra are referenced to $CDCl_3$ at 7.26 ppm for proton and 77.4 ppm for carbon. Coupling constants are given to 1 decimal place and are in Hertz (Hz). The abbreviations used to denote ^{1}H NMR are as follows; s, singlet; d doublet; dd, doublet of doublets; dq, doublet of quartets; t, triplet; q, quartet; qd, quartet of doublets; pent., pentet; m, multiplet; br, broad. Abbreviations used to describe ^{1}H NMR assignment are as follows; Ar, aryl. Signal assignment was aided by analysis of COSY, HMBC and HMQC where necessary.

TLC

Thin layer chromatography was carried out on Merck glass backed TLC plates coated with 0.2 mm silica gel or Merck DC-Alufolien Kieselgel 60 F254 aluminium-backed silica plates TLC plates. Plates were visualised with 254 nm UV light and developed with either a $KMnO_4$ or *p*-anisaldehyde solution and developed with a heat gun, where appropriate.

IR

Infrared (IR) spectroscopy was recorded on a Perkin Elmer Spectrum 100 FT-IR spectrometer with an ATR diamond cell irradiating between 4000 cm^{-1} and 600 cm^{-1}.

Melting point

Melting points were determined using an electrothermal melting point apparatus and are uncorrected.

Optical rotation

Optical rotations were determined with the sodium D line (λ= 589 nm) using a Perkin Elmer 241 MC polarimeter. $[\alpha]_D^{21}$ values are quoted in units 10^{-1} deg cm^2 g^{-1}.

Mass spectra

Mass spectra (both high and low resolution) were recorded on a Bruker Daltonics Apek 4e 7.0T FT-MS mass spectrometer mass and were obtained by electrospray ionisation (ESI) or chemical ionisation (CI).

LCMS

Chromatography was performed using a Water 2695 HPLC system and phenomenex Kinetic$_{TM}$ C_{18} reverse-phase column (150 x 2.5 mm, 100 Å with a flow rate of 1 mL min^{-1}. Solvents were A: H_2O + 0.05% formic acid and B: CH_3CN + 0.05% formic acid. A gradient elution was performed as follows over 10 minutes:

Time (minute)	%B
0.00	10
6.00	58
6.50	90
7.50	90
8.50	10
10.00	10

Single Ion Response (SIR) of certain mass ions was monitored over different time points as follows:

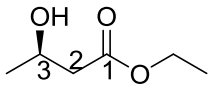
Event	Mass ion	Time (minutes)
MS Scan	100-600 ES-	0.00-10.00
SIR (internal standard)	196.40	1.40-3.40
SIR (Substrate)	242.20	2.00-4.00
SIR (dehydrated product)	224.20	4.00-6.00

Typically, 20 µL of sample was injected and detection was by Waters 2998 PDA and Waters Quatro micro API MS. MS was performed in ESI+ and ESI- modes. Typical MS conditions were as follows: capillary voltage (kV) 3.3; cone voltage (V): 30, desolvation gas flow (L min^{-1}): 600; cone gas flow (L min^{-1}): 50.

7.8 Synthesis of diketide substrates

7.8.1 (3*R*,2*R*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 114

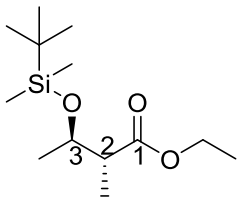
7.8.1.1 Ethyl (*R*)-hydroxybutanoate[109] 141



141

A solution of poly-3-(*R*)-hydroxybutyrate (10 g) in ethanol (70 mL), 1,2-dichloroethane (60 mL) and concentrated sulfuric acid (4 mL) was heated to reflux for 96 hours. The mixture was allowed to cool to room temperature and filtered through celite. The celite was washed with dichloromethane (100 mL) and the filtrate washed with $NaHCO_3$ (100 mL), brine (100 mL), dried ($MgSO_4$) and concentrated *in vacuo*. The crude material was distilled (72 °C at 11 mmHg) to give the title compound **141** (colourless oil, 8.40 g, 63 mmol, 55%). $[\alpha]_D^{22}$ -38.0 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -44 (c 1.0, $CHCl_3$)][109]; δ_H (400 MHz, $CDCl_3$): 1.12 (3H, d, *J* = 6.3 Hz, 3-CH_3), 1.26 (3H, t, *J* = 7.2 Hz, $OCH_2\underline{CH_3}$), 2.44 (2H, qd, *J* = 8.7, 3.6 Hz, 2-CH_2), 3.05 (1H, br s, OH), 4.16 (3H, m, OCH_2 & 3-CH); δ_C (400 MHz, $CDCl_3$); 14.2 ($OCH_2\underline{CH_3}$), 22.5 (3-CH_3), 42.9 (2-CH_2), 60.7 (3-CH), 64.4 (OCH_2), 172.9 (1-CO); $[\alpha]_D$, ^1H and ^{13}C NMR data in accordance with the literature.[109]

7.8.1.2 Ethyl (2*R*,3*R*)-3-(tert-butyldimethylsilyloxy)-2-methylbutanoate[111] 143

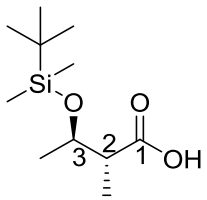


143

Diisopropylamine (12.2 mL, 87.1 mmol) was dissolved in anhydrous THF (80 mL) under nitrogen and cooled to -78 °C. *n*-Butyl lithium (2.36 M, 27.1 mL, 87.1 mmol) was added dropwise and the mixture stirred at -78 °C for 1 hour. A solution of ethyl (3*R*)-3-hydroxybutyrate **141** (5.01 g, 37.9 mmol) in anhydrous THF (20 mL) was added dropwise. The cold bath was removed and the mixture stirred at room temperature for 20 minutes, after which it was cooled back to -78 °C and methyl iodide (1.4 eq, 3.30 mL, 53.0 mmol) was added dropwise. The reaction was stirred at 0 °C for 3 hours, then quenched with 6 M aq. HCl (60 mL). The layers were separated and the aqueous layer was extracted with Et_2O (4 x 40 mL). The organic phases were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*.

The resulting residue was dissolved in anhydrous CH_2Cl_2 (100 mL) under nitrogen. Pyridine (6.40 mL, 78.5 mmol) was added. The mixture was cooled to 0 °C and *tert*-butyldimethylsilyl triflate (10.4 mL, 45.4 mmol) was added dropwise. The reaction was stirred at room temperature for 1 hour, after which it was quenched with sat. aq. $NaHCO_3$, (20 mL). The layers were separated and the aqueous layer was extracted with CH_2Cl_2 (2 x 100 mL). The organic phases were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification of the crude residue by flash chromatography (5% EtOAc, in PE) gave the title compound **143** (4.8 g, 32.9 mmol, 86% over two steps) as a colourless oil. $[\alpha]_D^{21.2}$ - 36.0 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -31.5 (c 1.9, $CHCl_3$)][116]; δ_H (400 MHz, $CDCl_3$); 0.02 (3H, s, $SiCH_3$), 0.05 (3H, s, $SiCH_3$), 0.85 (9H, s, 3 x CH_3), 1.07 (3H, d, *J* = 7.1 Hz, 2-CH_3), 1.10 (3H, d, *J* = 6.2 Hz, 3-CH_3), 1.25 (3H, t, *J* = 6.9 Hz, $OCH_2\underline{CH_3}$), 2.47 (1H, pent.. *J* = 7.3 Hz, 2-H), 4.01 (1H, dq, *J* = 6.4, 6.2 Hz, 3-H), 4.10 (2H, q, *J* = 7.1 Hz, OCH_2); δ_C (400 MHz, $CDCl_3$); -4.9 (Si-CH_3), -4.23 (Si-CH_3), 12.8 (2-CH_3), 14.3 ($OCH_2\underline{CH_3}$), 18.1 (Si-C(CH_3)$_3$) 20.7 (3-CH_3), 25.9 (Si-C(CH_3), 48.3 (2-CH), 60.3 (OCH_2), 70.3 (3-CH), 175.3 (CO). ^1H and ^{13}C NMR data in accordance with the literature.[111]

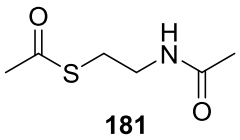
7.8.1.3 (2*R*, 3*R*)-3-(Tert-butyldimethylsilyloxy)-2-methylbutanoic acid[170] 144



144

MeOH (12 mL) and lithium hydroxide (414 mg, 17.2 mmol) in water (2 mL) were added to ester **143** (1.50 g, 5.76 mmol) dissolved in THF (5 mL). The reaction mixture was heated to 60 °C overnight. The reaction was diluted with water (8 mL) and extracted with Et_2O (2 x 15 mL). The aqueous layer was acidified to pH 3 with H_2SO_4 (6 M) and further extracted with EtOAc (3 x 15 mL), the pH was adjusted to pH 3 as required. The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo* to give a yellow oil. Purification by flash chromatography (SiO_2, 10% EtOAc/ Petrol/ 0.5% acetic acid, visualised with $KMnO_4$) afforded the title compound **144** (550 g, 2.37 mmol, 41%) as a colourless oil. $[\alpha]_D^{23.0}$ -7.0 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -12.5 (c 1.0, $CHCl_3$)][170]; δ_H (400 MHz, $CDCl_3$); 0.11 (3H, s, $SiCH_3$), 0.13 (3H, s, $SiCH_3$), 0.91 (9H, s, $SiC(CH_3)_3$), 1.23 (3H, d, *J* = 7.0 Hz, 2-CH_3), 1.25 (3H, d, *J* = 6.1 Hz, 3-CH), 2.52 (1H, dq, *J* = 7.5, 4.5 Hz, 2-CH) 3.97 (1H, q, *J* = 7.2 Hz, 3-CH); δ_C (400 MHz, $CDCl_3$); 5.0 ($SiCH_3$), 4.2 ($SiCH_3$), 13.3 (2-CH_3), 18.0 ($SiC(CH_3)_3$) 21.0 (3-CH_3), 25.8 ($SiC(CH_3)_3$), 48.0 (2-CH), 70.4 (3-CH), 175.4 (CO). ν_{max}(neat) / cm^{-1}:2930 (OH), 2887 (CH), 1708 (C=O), 1115 (CO), 831 (Si-CH_3). ^{1}H NMR and IR data in accordance with the literature.[170, 171]

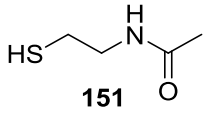
7.8.1.4 *N,S*-diacetylcysteamine[172] 181



181

Potassium hydroxide (6M) was added to a solution of cysteamine hydrochloride (30.8 mmol, 3.60 g) in water (20 mL) at 0 °C until pH 8 was reached. Acetic anhydride (92.4 mmol, 5.82 mL) was added dropwise keeping the reaction at pH 8 using potassium hydroxide (6 M). After the addition of acetic anhydride the pH was adjusted to 7 with hydrochloric acid (2 M) and stirred at 0 °C for 1.5 hours. The mixture was extracted with CH_2Cl_2 (5 x 25 mL), dried ($MgSO_4$), filtered and concentrated in *vacuo* to give the title compound **181** as a colourless oil, (2.70 g, 54% yield). δ_H (400 MHz, $CDCl_3$): 1.96 (3H, s, CH_3), 2.34 (3H, s, CH_3), 3.00 (2H, t, J = 6.5 Hz, S-CH_2), 3.41 (2H, dt, J = 6.5 Hz, N-CH_2), 6.06 (1H, br s, NH); δ_c (400 MHz, $CDCl_3$): 23.3 (4-CH_3), 28.9 (1-CH_3), 30.8 (S-CH_2), 39.7 (N-CH_2), 170.6 (1-CO), 196.4 (4-CO); ν_{max} (oil)/cm-^{1}: 3284 (N-H), 2932 (C-H), 1689 (C=O), 1650 (C=O), 1431 (C-CH_3), 1356 (C-CH_3), 1132 (C-S). ^{1}H and ^{13}C NMR data in accordance with the literature.[111]

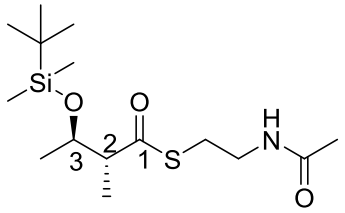
7.8.1.5 *N*-acetylcysteamine[172] 151



151

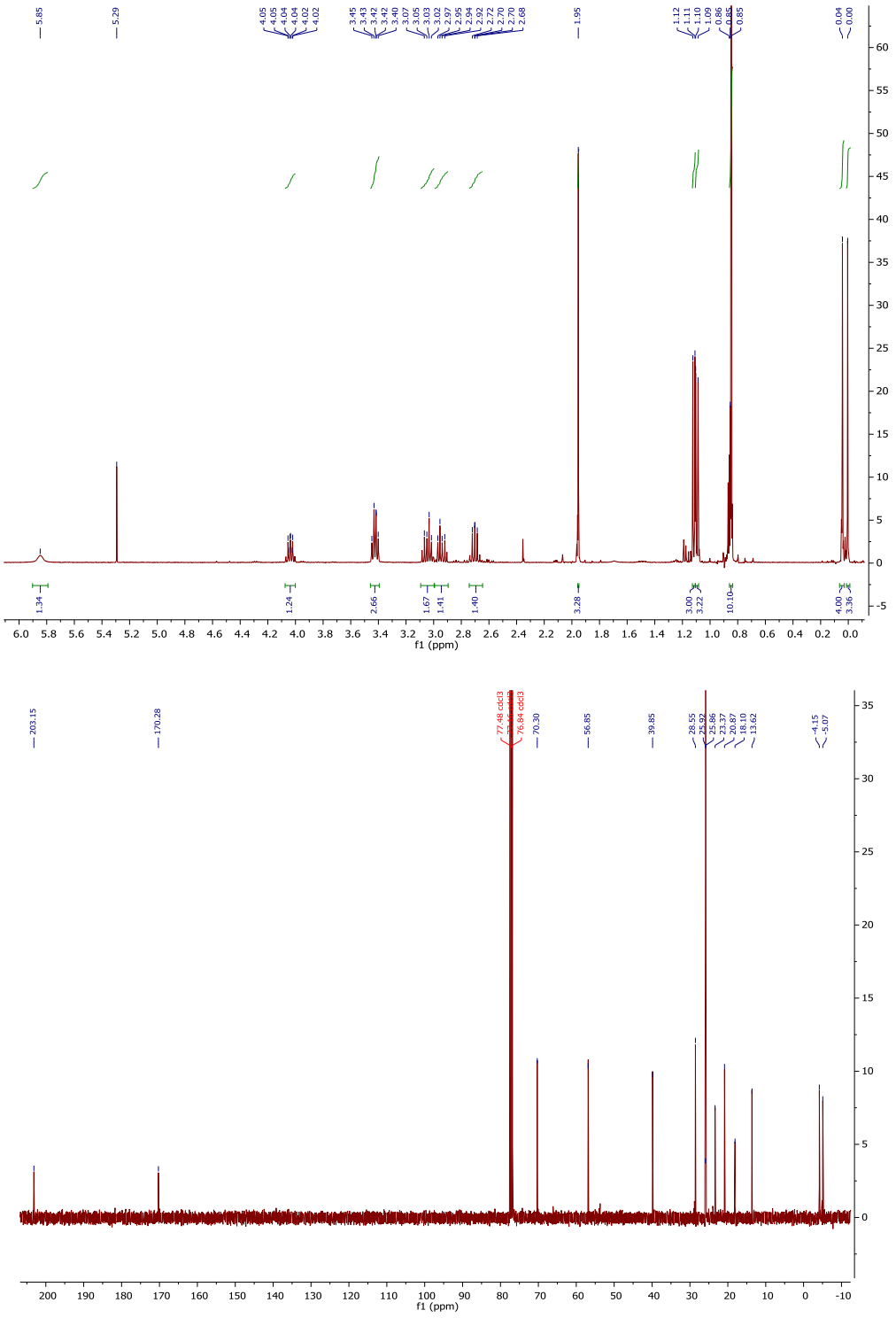
Potassium hydroxide (9.07 g, 155 mmol) was added to a solution of *N,S*-diacetylcysteamine **151** (5.28 g, 31.0 mmol) dissolved in water (30 mL) at 0 °C. The reaction was warmed to room temperature and stirred for 1 hour before being cooled to 0 °C and acidified to pH 5. The aqueous layer extracted with CH_2Cl_2 (3 x 30 mL). The organic extracts were combined, dried ($MgSO_4$) and concentrated *in vacuo* to give the title compound **151** as a colourless oil (3.01 g, 26.0 mmol, 82% yield). δ_H (400 MHz, $CDCl_3$): 1.36 (1H, t, J = 8.1 Hz, SH), 1.83 (3H, s, 1-CH_3), 2.48 (2H, q, J = 7.0 Hz, CH_2S), 3.21 (2H, q, J = 6.6 Hz, CH_2N); δ_C (400 MHz, $CDCl_3$)[173]: 22.6 (1-CH_3), 23.9 (3-CH_2), 42.4 (2-CH_2), 170.8 (1-CO); ^{1}H and ^{13}C NMR data in accordance with the literature.[172, 173]

7.8.1.6 (2*R*, 3*R*)-3-Tert-butyldimethylsilyloxy-2-methylbutanoyl-*N*-actylcysteamine 145



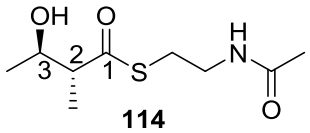
145

EDCI (405 mg, 2.60 mmol) and DMAP (23.0 mg, 0.20 mmol) were added to a solution of acid **144** (402 mg, 1.80 mmol) in anhydrous CH_2Cl_2 (25 mL) at 0 °C under nitrogen . After stirring at 0 °C for 15 minutes, HSNAC **151** was added in anhydrous CH_2Cl_2 (10 mL). The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (15 mL) and the layers separated. The aqueous phase was extracted with CH_2Cl_2 (3 x 25 mL). The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography gave the title compound **145** (360 mg, 1.10 mmol, 59%) as a pale yellow oil. $[\alpha]_D^{21.9}$ -69.4 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 0.00 (3H, s, Si-CH_3), 0.04 (3H, s, Si-CH_3), 0.85 (9H, s, Si-C(CH_3)$_3$), 1.10 (3H, d, *J* = 7.0 Hz, 2-CH_3), 1.12 (3H, d, *J* = 6.1 Hz, 3-CH_3), 2.70 (1H, dt. *J* = 7.4 Hz, 2-CH), 2.99 (2H, m, S-CH_2), 3.42 (2H, m, N-CH_2), 4.04 (1H, dq, *J* = 7.2 Hz , 3-CH), 5.85 (1H, br s, NH); δ_C (400 MHz, $CDCl_3$); -5.1 (Si-CH_3), -4.2 (Si-CH_3), 13.6 (2-CH_3), 18.1 (Si-$\underline{C}$(CH_3)$_3$), 20.9 (3-CH_3), 23.4 (CH_3), 25.9 (Si-C($\underline{C}$$H_3$)$_3$), 28.6 (S-$CH_2$), 38.9 (N-$CH_2$), 56.9 (2-CH), 70.3 (3-CH), 170.3 (N-CO), 203.2 (S-CO); ν_{max} (neat) / cm^{-1}: 3284 (NH), 2955 (CH), 2929 (CH),2856 (CO-CH_3), 1688 (S-CO), 1654 (N-CO), 836 (Si-CH_3); *m/z* (ESI) 356.17 [M+Na^+]; Found (ESI) 356.1686 ($C_{15}H_{31}NaNO_3SSi$ requires 356.1678).



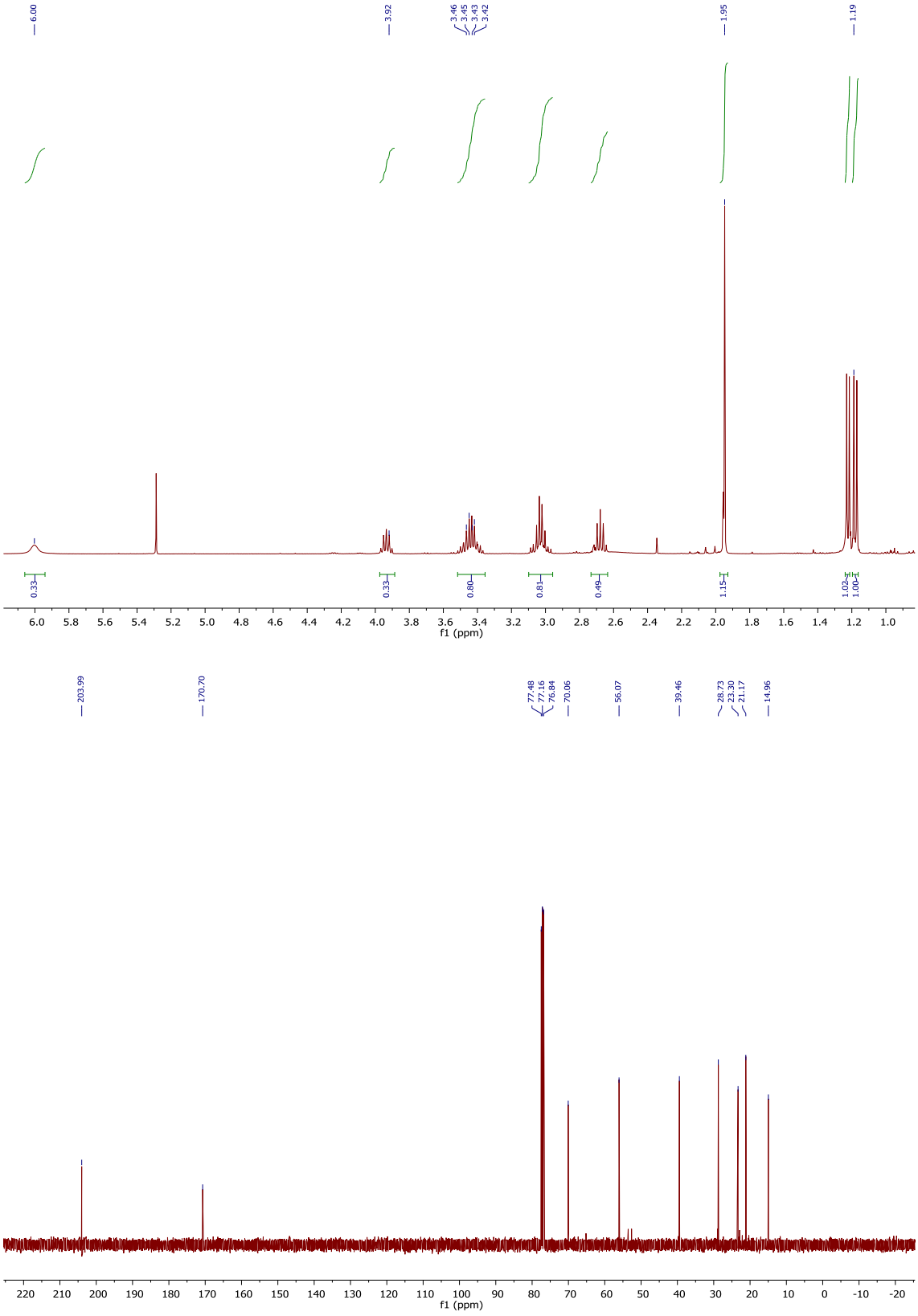
^{1}H, 400 MHz and ^{13}C NMR, 100 MHz, of (3*R*, 2*R*) **145** in $CDCl_3$.

7.8.1.7 (2*R*,3*R*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 114



114

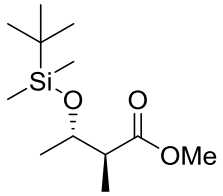
TBDMS-ether **145** (150 mg, 0,45 mmol) was dissolved in THF (1.3 mL), water (1.3 mL) and acetic acid (4 mL) and stirred at room temperature for 5 days. The reaction was diluted with water (25 mL) and the aqueous layer washed with CH_2Cl_2 (25 mL). The aqueous layer was concentrated to give the title compound **114** (78.0 mg, 0.35 mmol, 79%) as a colourless oil. $[\alpha]_D^{22.2}$ -32.8 (*c* 0.33, $CHCl_3$); δ_H (400 MHz, $CDCl_3$);1.19, (3H, d, *J* = 7.1 Hz, 2-CH_3), 1.23 (3H, d, *J* = 6.3 Hz, 3-CH_3), 1.96 (3H, s, CO-CH_3),2.69 (1H, dt, *J* = 7.2 Hz, 2-CH), 3.05 (2H, m, S-CH_2), 3.45 (2H, m, N-CH_2), 3.94 (1H, dt, *J* = 6.2 Hz, 3-CH), 5.88 (1H, br s, NH); δ_C (400 MHz, $CDCl_3$); 14.9 (2-CH_2), 21.2 (3-CH_3), 23.3 (CO-$\underline{C}H_3$), 28.7 (S-CH_2) 39.5 (N-CH_2), 56.1 (2-CH), 70.1 (3-CH), 170.7 (N-C=O), 204.0 (S-C=O);); ν_{max} (neat) / cm^{-1}: 3287 (NH), 3134 (OH), 2973 (CH), 2932 (CH), 1654 (NC=O), 1548 (SC=O); 1438 (NH); *m/z* (ESI) 242.08 [M+Na^+]; Found (ESI) 242.0822 ($C_9H_{17}NaNO_3S$ requires 242.0821). ^{1}H NMR data in accordance with the literature.[99]



^{1}H 400 MHz and ^{13}C NMR, 100 MHz, of (3*R*, 2*R*) **114** in $CDCl_3$.

7.8.2 (2*S*,3*S*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 115

7.8.2.1 Methyl (2*S*, 3*S*)-3-(*tert*-butyldimethylsilyloxy)-2-methylbutanoate 154[118]

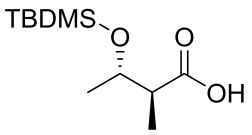


154

Diisopropylamine (5.0 mL, 35.5 mmol) was dissolved in anhydrous THF (30 mL) under nitrogen and cooled to -78 °C. *n*-Butyl lithium (2.15 M, 16.5 mL, 35.5 mmol) was added dropwise and the mixture stirred at -78 °C for 1 hour. A solution of methyl (3*S*)-3-hydroxybutyrate (2.00 g, 16.5 mmol) in anhydrous THF (12 mL) was added dropwise. The cold bath was removed and the mixture stirred at room temperature for 20 minutes, after which it was cooled back to -78 °C and methyl iodide (1.30 mL, 21.1 mmol) was added dropwise. The reaction was stirred at 0 °C for 3 hours, then quenched with 6 M aq. HCl (20 mL). The layers were separated and the aqueous layer was extracted with Et_2O (4 x 20 mL). The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*.

The resulting residue was dissolved in anhydrous CH_2Cl_2 (40 mL) under nitrogen. Pyridine (2.90 mL, 35.5 mmol) was added. The mixture was cooled to 0 °C and *tert*-butyldimethylsilyl triflate (4.70 mL, 20.3 mmol) was added dropwise. The reaction was stirred at room temperature for 1 hour, after which it was quenched with sat. aq. $NaHCO_3$, (20 mL). The layers were separated and the aqueous layer was extracted with CH_2Cl_2 (2 x 40 mL). The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification of the crude residue by flash chromatography (5% EtOAc, in petroleum ether) gave the title compound **154** (2.50 g, 10.3 mmol, 63% over two steps) as a pale yellow oil. $[\alpha]_D^{22.5}$ +37.3 (*c* 1.0, $CHCl_3$), [lit $[\alpha]_D$ +37.1 (*c* 1.43, CH,Cl_3)][118] δ_H (400 MHz, $CDCl_3$); 0.01 (3H, s, Si-CH_3), 0.04 (3H, s, Si-CH_3), 0.85 (9H, s, Si-C(CH_3)$_3$), 1.06 (3H, d, *J* = 7.1 Hz, 2-CH_3), 1.10 (3H, d, *J* = 6.2 Hz, 3-CH_3), 2.49 (1H, dq., *J* = 7.2 Hz, 2-CH), 3.65 (3H, s, CO-CH_3), 3.99 (1H, dq. *J* = 6.2, 7.4 Hz, 3-CH). δ_C (400 MHz, $CDCl_3$); -5.0 (Si-CH_3), -4.2 (Si-CH_3), 12.9 (2-CH_3), 18.0 (Si-$\underline{C}$(CH_3)), 20.7 (3-CH_3), 25.8 (Si-C($\underline{CH_3}$)$_3$), 48.3 (2-CH), 51.5 (O-CH_3), 70.3 (3-CH), 175.3 (CO); ν_{max}(neat) / cm^{-1}: 2953 (CH), 2930 (CH), 2857 (CH), 1740 (C=O), 1252 (CO), 833 (Si-CH_3). ^1H and ^{13}C NMR data in accordance with the literature.[117, 118]

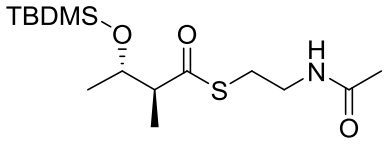
7.8.2.2 (2*S*, 3*S*)-3-(*Tert*-butyldimethylsilyloxy)-2-methylbutanoic acid 155[174]



155

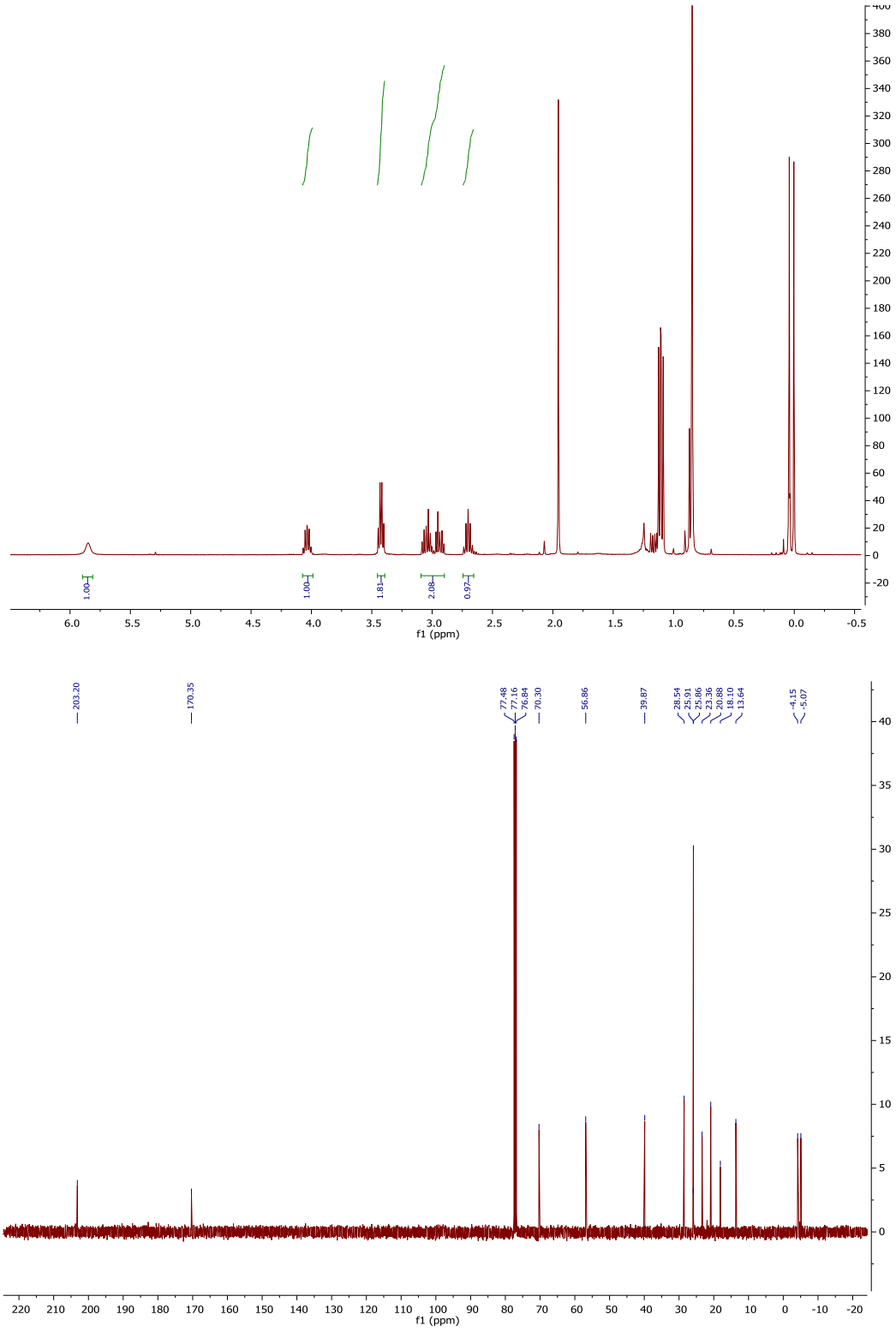
Ester **154** (1.00 g, 4.1 mmol) was dissolved in THF (5 mL) and MeOH (8 mL) and lithium hydroxide (291 mg, 12.2 mmol) in water (1.5 mL) was added. The reaction mixture was heated to 60 °C overnight. The reaction was cooled and diluted with water (7 mL) and extracted with Et_2O (2 x 10 mL). The aqueous layer was acidified to pH 3 with H_2SO_4 (6 M) and further extracted with EtOAc (3 x 15 mL), the pH was adjusted to pH 3 as required. The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo* to give a yellow oil. Purification by flash chromatography (SiO_2, 10% EtOAc/ petroleum ether, 0.5% acetic acid visualised with $KMnO_4$) afforded the title compound **155** (416 g, 1.79 mmol, 44%) as a pale yellow oil. $[\alpha]_D^{22.6}$ +7.9 (*c* 1.0, $CHCl_3$), [lit $[\alpha]_D$ +5.7 (*c* 0.88, CH,Cl_3)][174]; δ_H (400 MHz, $CDCl_3$); 0.07 (3H, s, Si-CH_3), 0.08 (3H, s, Si-CH_3), 0.88 (Si-C(CH_3)$_3$), 1.16 (3H, d, *J* = 7.1 Hz, 2-CH_3), 1.18 (3H, d, *J* = 6.2 Hz, 3-CH_3), 2.51 (1H, dq., *J* = 6.1 Hz, 2-CH), 4.00 (1H,dq , *J* = 6.2 Hz, 3-CH); ν_{max} (neat) / cm^{-1}: 2953 (CH), 2930 (CH), 2886 (CH), 2857 (CH), 1740 (CO). ^1H NMR data in accordance with the literature.[174]

7.8.2.3 (2*S*, 3*S*)-3-(*Tert*-butyldimethylsilyloxy)-2-methylbutanoyl-*N*-acetylcysteamine 156



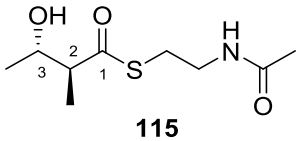
156

Acid **155** (225 mg, 0.92 mmol) was dissolved in anhydrous CH_2Cl_2 (10 mL) at 0 °C under nitrogen. EDCI (245 mg, 1.28 mmol) and DMAP (1.0 mg, 0.10 mmol) were added. After stirring at 0 °C for 15 minutes, HSNAC **151** (152 mg, 1.28 mmol) was added in dry CH_2Cl_2 (10 mL). The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (5 mL) and the layers separated. The aqueous phase was extracted with CH_2Cl_2 (3 x 15 mL). The organics were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 3% MeOH/ CH_2Cl_2 visualised with $KMnO_4$) gave the title compound **156** (169 mg, 0.51 mmol, 55%) as a pale yellow oil. $[\alpha]_D^{22.7}$ +56.9 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 0.01 (3H, s, Si-CH_3), 0.05 (3H, s, Si-CH_3), 0.85 (9H, s, Si-C$(CH_3)_3$), 1.10 (3H, d, *J* = 7.0 Hz, 2-CH_3), 1.12 (3H, d, *J* = 6.1 Hz, 3-CH_3), 1.96 (3H, s, CO-CH_3), 2.71 (1H, dq *J* = 7.2 Hz, 2-CH), 3.00 (2H, m, S-CH_2), 3.43 (2H, m, N-CH_2), 4.03 (1H, dq. *J* = 6.8 Hz, 3-CH); δ_c (400 MHz, $CDCl_3$); -5.1 (Si-CH_3), -4.2 (Si-CH_3), 13.6 (2-CH_3), 18.1 (Si$\underline{C}(CH_3)_3$), 20.9 (3-CH_3), 23.4 (CO-CH_3), 25.9 (SiC($\underline{C}H_3)_3$), 28.5 (S-CH_2), 39.9 (N-CH_2), 56.9 (2-CH), 70.3 (3-CH), 170.0 (N-CO), 203.2 (S-CO); ν_{max} (neat) / cm^{-1}: 3285 (NH), 2955 (CH), 2929 (CH), 2884 (CH), 2586 (CH), 1687 (CO), 1653 (CO);); *m/z* (ESI) 356.17 [M+Na^+]; Found (ESI) 356.1683 ($C_9H_{17}NaNO_3S$ requires 356.1683).



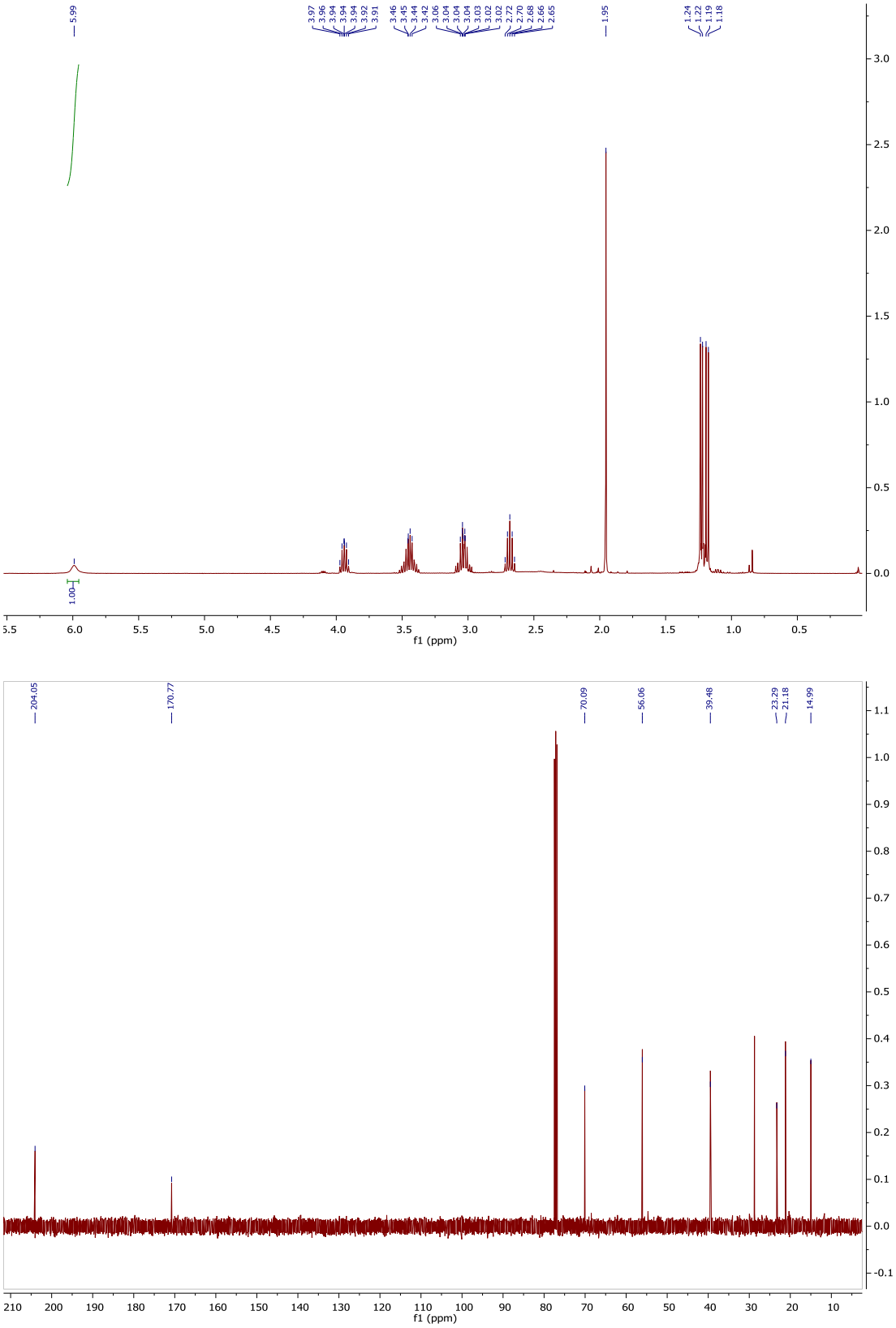
^{1}H, 400 HHz, and ^{13}C NMR, 100 MHz, of (3*S*, 2*S*) **156** in $CDCl_3$.

7.8.2.4 (3*S*, 2*S*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 115



115

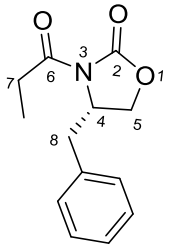
TBDMS-ether **156** (100 mg, 0.30 mmol) was dissolved in THF (0.9 mL), water (0.9 mL) and acetic acid (2.6 mL) and stirred at room temperature for 5 days. The reaction was diluted with water (25 mL) and the aqueous layer washed with CH_2Cl_2 (25 mL). The aqueous layer was concentrated to give the title compound **115** (37 mg, 0.16 mmol, 56%) as a colourless oil. $[\alpha]_D^{23.9}$ +36.8 (*c* 0.64, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 1.17 (3H, d, *J* = 7.1 Hz, 2-CH_2), 1.23 (3H, d, *J* = 6.3 Hz, 3-CH_3), 1.95 (3H, s, CO-CH_3), 2.68 (1H, pent. *J* = 7.2 Hz, 2-CH), 3.02 (2H, m, S-CH_2), 3.43 (2H, m, N-CH_2), 3.94 (1H, pent., *J* = 6.8 Hz, 3-CH), 6.15 (1H, br s, NH); (400 MHz, $CDCl_3$); 15.0 (2-CH_3), 21.2 (3-CH_3), 23.3 (NCO-CH_3), 28.7 (S-CH_2), 39.5 (N-CH_2), 56.1 (2-CH), 70.1 (3-CH), 170.8 (NC=O), 204.1 (SC=O); ν_{max}(neat) / cm^{-1}:3287 (NH), 3134 (OH), 2973 (CH), 2931 (CH), 1653 (NC=O), 1549 (SC=O), 1449 (NH);); *m/z* (ESI) 242.08 [M+Na^+]; Found (ESI) 242.0819 ($C_9H_{17}NaNO_3S$ requires 242.0821).



^{1}H and ^{13}C NMR, 400 MHz, of (3*S*, 2*S*) **115** in $CDCl_3$.

7.8.3 (2*S*,3*R*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine

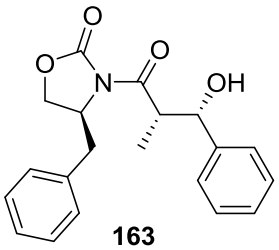
7.8.3.1 (*S*)-4-Benzyl-3-propionyloxazolidine-2-one[175] 129



129

(*S*)-4-Benzyl-2-oxazolidinone **127** (1.01 g, 5.60 mmol) was dissolved in dry THF (20 mL) under N_2 and cooled to -78 °C, n-BuLi (1.95 M in hexane) (2.80 mL, 5.60 mmol) was added dropwise and the solution stirred for 20 minutes. Propionyl chloride (0.60 mL, 6.20 mmol) was added and stirred for 3 hours before being warmed to room temperature. The reaction was quenched with NH_4Cl aq. (5 mL) and the majority of the THF was removed *in vacuo*. CH_2Cl_2 (10 mL) was added and the solution washed with 10% aq NaOH. The aqueous layer was extracted with CH_2Cl_2 (2 x 20 ml). The organic layers were combined, washed (brine), dried ($MgSO_4$), filtered and concentrated *in vacuo* to give the title compound **129** (0.93 g, 3.98 mmol, 72% yield) as a white solid. mp 41.5-43 °C (from CH_2Cl_2), $[\alpha]_D^{24.3}$ +82.5 (*c* 0.3, $CHCl_3$), [lit $[\alpha]_D$ +90.4 (c = 0.95, $CHCl_3$)][176]; δ_H (400 MHz, $CDCl_3$); 1.21 (3H, t, *J* = 7.4 Hz, 7-CH_3), 2.77 (1H, dd, *J* = 13.4, 9.6 Hz, 8-CH_2), 2.96 (2H, q, *J* = 7.3 Hz, 7-CH_2), 3.31 (1H, dd, *J* = 13.4, 3.4 Hz, 8-CH_2), 4.19 (2H, m, 5-CH_2O), 4.67 (1H, m, 4-CHN), 7.29 (5H, m, Ar-H); ν_{max} (neat) / cm^{-1}: 2983 (CH), 2940 (CH), 2919 (CH), 1780 (CO), 1699 (CO). ^1H NMR data in accordance with the literature.[175]

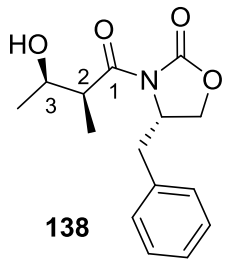
7.8.3.2 (4*S*)-3-[(2'*S*,3'*R*)-3'-hydroxy-2'-methyl-3'phenylpropanoyl]-4-phenylmethyl-2-oxazolidinone[102] 163



163

(*S*)-4-benzyl-3-propionyloxazolidin-2-one **129** (1 eq, 207 mg, 0.858 mmol) was dissolved in dry $CHCl_2$ (2 mL) under N_2, cooled to 0 °C and gently deoxygenated. Dibutyl boron triflate (1.2 eq, 1.03 mL, 1.03 mmol) was added slowly dropwise. Triethylamine (1.2 eq, 0.14 mL, 1.03 mmol) was then added slowly dropwise before the solution was cooled to -78 °C and acetaldehyde (10 eq, 0.5 mL 8.89 mmol) added. The solution was stirred at -78 °C for 20 minutes, before being allowed to warm slowly to 0 °C over 3.5 hours. The reaction was quenched with MeOH/1M phosphate buffer (8 mL:3:1). MeOH/30 % aq. hydrogen peroxide (6 mL, 2:1) was then added slowly, keeping the temperature below 10 °C and stirred for 1 hour. The volatiles were removed and the resulting slurry extracted with ether (3 x 10 mL). The ether layers were combined and washed with 5 % $2NaHCO_3$ (6 mL, 0.3 g) and brine (6 mL). Sodium sulfite solution (2M) (6 mL) was then added to the ether layer and stirred for 1 hour before being extracted with ether (2 x 10 mL). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo*. The crude product was purified using column chromatography (solvent system 50:50 ethyl acetate: petrol (40-60 °C) to give the title compound Figure 40) **163** (clear colourless oil, 149 mg, 0.594 mmol, 59 %). $[\alpha]_D^{22}$ +47.5 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ + 70.4 (c 1.04, $CHCl_3$)][102]; δH (400 MHz, $CDCl_3$); 1.23 (3H, d, *J* = 6.9 Hz, 2-CH_3), 2.77 (1H, dd, *J* = 13.4, 9.5 Hz, CH*2*Ph), 3.05 (1H, br s, OH), 3.25 (1H, dd, *J* = 13.5, 3.4 Hz, CH_2Ph), 4.10 (3H, m, CH_2O & CHN), 4.60 (1H, m, 2-CH), 5.10 (1H, d, *J* = 3.7 Hz, 3-CH), 7.32 (10H, m, Ar-H). δc (400 MHz, $CDCl_3$); 11.1 (2-CH_3), 38.0 (CH_2Ph), 55.68 (CHN), 66.43 (CH_2O), 55.1 (2-CH), 73.83 (3-CH), 126.3 (Ar-C), 127.6 (Ar-C), 127.7 (Ar-C), 128.4 (Ar-C), 129.1 (Ar-C), 129.6 (Ar-C).

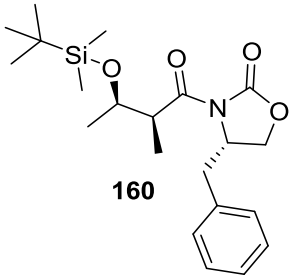
7.8.3.3 (4*S*)-3-[(2*S*,3*R*)-3-Hydroxy-2-methylbutyl]-4-phenylmethyl-2-oxazolidinone **138**[102]



138

(*S*)-4-Benzyl-3-propionyloxazolidin-2-one **129** (850 mg, 3.65 mmol) was dissolved in anhydrous CH_2Cl_2 (10 mL) and cooled to -78 °C under N_2. Dibutyl boron triflate (1 M in CH_2Cl_2, 7.30 mL, 7.30 mmol) was added dropwise over 1 hour. After 10 minutes triethylamine (1.27 mL, 9.13 mmol) was added dropwise and the solution stirred for 1 hour at -78 °C, then 100 minutes at 0 °C. The solution was cooled to -78 °C and a solution of acetaldehyde (10 eq, 2.05 mL, 36.5 mmol) in anhydrous CH_2Cl_2 (1.0 mL) also at -78 °C was added slowly. After stirring for 2 hours at -78 °C and then 100 minutes at 0 °C, phosphate buffer (pH 7.1 M, 28 mL) and MeOH (28 mL) were added, the resulting mixture was stirred vigorously in an ice bath as 30% H_2O_2 (14 mL) was added dropwise. After 30 minutes at 10-15 °C, the solvent was removed *in vacuo* and the resulting residue was partitioned between saturated $NaCO_3$ aq. and CH_2Cl_2 (8 mL). The aqueous layer was extracted with CH_2Cl_2 (2 x 8 mL) and the combined organic layers were washed with saturated $NaCO_3$ aq. (8 mL), dried ($MgSO_4$), filtered, and concentrated *in vacuo* to give a crude pale yellow oil. The crude material was purified by flash chromatography (SiO_2, 20% to 40% EtOAc /petroleum ether. visualized with $KMnO_4$) to give the aldol adduct **138** as a colourless oil (602 mg, 2.10 mmol, 69% yield). $[\alpha]_D^{22.0}$ +48.8 (*c* = 1.0, $CHCl_3$), [lit. $[\alpha]_D$ +51.1 (*c* = 1.0, $CHCl_3$)][121];δ_H (400 MHz, $CDCl_3$); 1.19 (3H, d, *J* = 6.4 Hz, 2-CH_3), 1.27 (3H, d, *J* = 7.0 Hz, 3-CH_3), 2.79 (1H, dd, *J* = 13.4, 9.4 Hz, CH_2Ph), 2.89 (1H, br s, OH), 3.26 (1H, dd, *J* = 13.5, 3.4 Hz, CH_2Ph), 3.75 (1H, dq, *J* = 7.1, 2.9 Hz, 2-CH), 4.21 (3H, m, 3'-CH_2 & 3-CH), 4.71 (1H, m, 4'-CH), 7.32, (5H, m, Ar-H); δ_C (400 MHz, $CDCl_3$) 10.6 (3-CH_3), 19.8 (2-CH_3), 38.0 (5'-CH_2), 43.3 (2-CH), 55.2 (4'-CH), 66.3 (3'-CH), 67.8 (3-CH), 127.3 (Ar-C), 129.1 (Ar-C), 129.6 (Ar-C), 135.1 (Ar-C), 153.3 (1'-CO), 177.5 (1-CO); ν_{max} (neat) / cm^{-1}: 3439 (OH), 2987 (CH), 1775 (CO), 1688 (CO), 889 (Ar-CH).^{1}H NMR data in accordance with the literature.[102, 121]

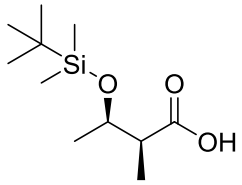
7.8.3.4 (4*S*)-3-[(2*S*,3*R*)- 3-*tert*-Butyldimethylsilyloxy-2-methylbutyl]-4-phenylmethyl-2-oxazolidinone 160 [121]



160

Aldol adduct **138** (1.00 g, 3.61 mmol) was dissolved in anhydrous CH_2Cl_2 (30 mL) under N_2. TBDMSCl (1.10 g, 7.20 mmol), imidazole (0.98 g, 14.4 mmol) and DMAP (3 crystals) were added and stirred overnight. The reaction was quenched with $NaHCO_3$ (34 mL), extracted with CH_2Cl_2. The organic extracts were dried ($MgSO_4$), filtered and concentrated. Purification by flash chromatography (SiO_2, 10% EtOAc in petroleum ether visualised with $KMnO_4$) gave **160** (532 mg, 1.36 mmol, 37%) as a colourless crystalline solid. $[\alpha]_D^{23.0}$ +51.8 (c = 1.0, $CHCl_3$), [lit. $[\alpha]_D$ +49.2 (c = 1.0, $CHCl_3$)][121] δ_H (400 MHz, $CDCl_3$): 0.02 (3H, s, $SiCH_3$), 0.05 (3H, s, $SiCH_3$), 0.87 (9H, s, $SiC(CH_3)_3$), 1.18 (3H, m, J = 6.1 Hz, 2-CH_3), 1.22 (3H, d, J = 6.8 Hz, 3-CH_3), 2.77 (1H, dd, J = 13.3, 9.7 Hz, CH_2Ph), 3.29 (1H, dd, J = 13.3, 3.3 Hz, CH_2Ph), 3.83 (1H, qd, J = 6.9, 5.7 Hz, 2-H), 4.08 (1H, m, J = 6.0 Hz, N-CH), 4.16 (2H, m, OCH_2), 4.63 (1H, m, 3-CH), 7.21-7.36 (5H, m, Ar-H); ν_{max} (neat)/cm^{-1}: 2955 (CH), 2929 (CH), 2885 (CH), 2856 (CH), 1778 (CO), 1694 (CO), 809 (Ar C); ^{1}H NMR data in accordance with the literature.[121]

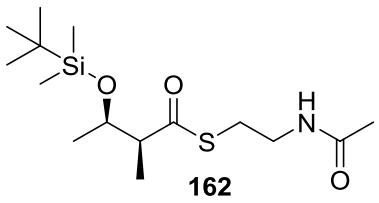
7.8.3.5 (2*S*,3*R*)-3-*tert*-Butyldimethylsilyloxy-2-methylbutanoic acid 161[174]



161

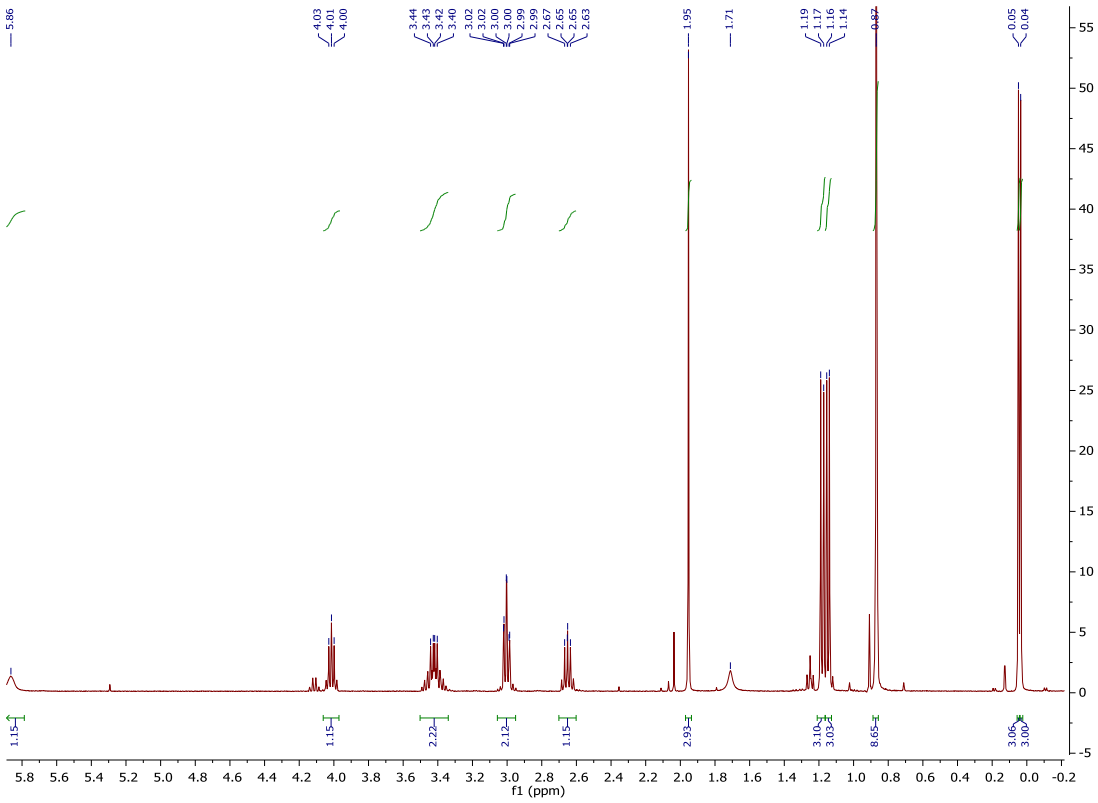
A solution of the protected aldol **160** (377 mg, 1.27 mmol) dissolved in THF/H_2O (20 mL/6 mL) was cooled to 0 °C. Hydrogen peroxide (6.35 mmol, 0.70 mL 30% w/v H_2O_2) and LiOH (61.0 mg 2.54 mmol) were added. The reaction was warmed to room temperature. After 4 hours the excess peroxide was quenched using 1.5 M sodium sulfate (Na_2SO_3) aq. and the THF removed *in vacuo*. The resulting slurry was acidified to pH 2 using 2M HCl and extracted with CH_2Cl_2 (3 x 20 mL). The organic extracts were combined, washed with brine (10 mL), dried, filtered and concentrated *in vacuo*. Purification of the crude by flash chromatography (SiO_2, 10 EtOAc/ petroleum ether/ 0.1% Acetic acid, visualised with $KMnO_4$) afforded the title compound **161** (181 mg, 0.78 mmol, 61%) as a colourless oil. $[\alpha]_D^{21.0}$ +20.6 (*c* 1.0, $CHCl_3$), [lit $[\alpha]_D$ +25 (c = 0.51, $CHCl_3$)][174]; δ_H (400 MHz, $CDCl_3$); 0.11 (3H, s, $SiCH_3$), 0.12 (3H, s, $SiCH_3$), 0.90 (9H, s, $SiC(CH_3)_3$, 1.14 (3H, d, *J* = 7.1 Hz, 2-CH_2), 1.18 (3H, d, *J* = 6.3 Hz, 3-CH_3), 2.58 (1H, qd, *J* = 4.9, 7.2 Hz, 2-CH), 4.10 (1H, m, 3-CH); ν_{max} (neat) / cm^{-1}: 2955 (CH), 2929 (CH), 2887 (CH), 2857 (CH), 1705 (CO). ^{1}H NMR data in accordance with the literature.[174]

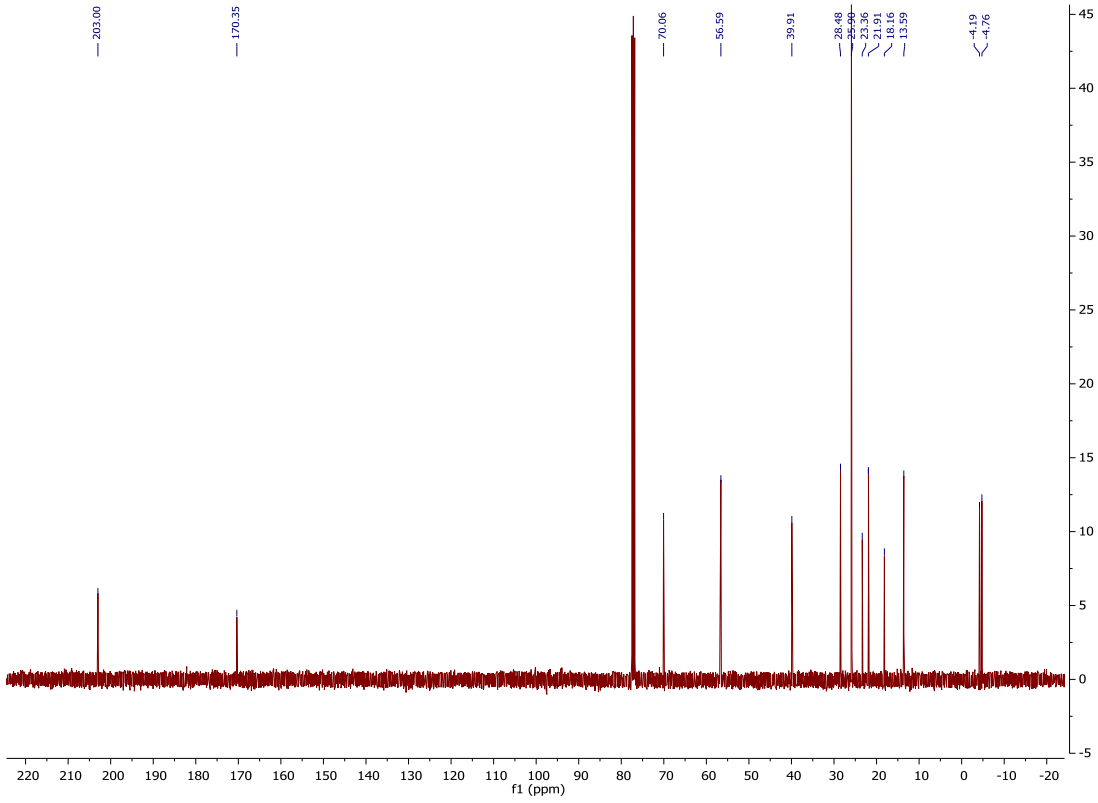
7.8.3.6 (2*S*,3*R*)- 3-*tert*-Butyldimethylsilyloxy-2-methylbutanoyl-*N*-acetylcysteamine **162**



162

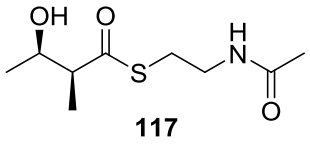
EDCI (93.0 mg, 0.60 mmol) and DMAP (5.00 mg, 0.04 mmol) were added to a solution of acid **161** (100 mg, 0.43 mmol) in anhydrous CH_2Cl_2 (6 mL) at 0 °C under nitrogen . After stirring at 0 °C for 15 minutes, HSNAC **151** (71.0 mg, 0.60 mmol) was added in anhydrous CH_2Cl_2 (2.5 mL). The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (7 mL) and the layers separated. The aqueous phase was extracted with CH_2Cl_2 (3 x 15 mL). The organics were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 3% MeOH/ CH_2Cl_2 visualised with $KMnO_4$) gave the title compound **162** (67.0 mg, 0.20 mmol, 46%) as a colourless oil. $[\alpha]_D^{22.0}$ +15.6 (*c* 1.0, $CHCl_3$), δ_H (400 MHz, $CDCl_3$); 0.04 (3H, s, Si-CH_3), 0.05 (3H, s, Si-CH_3), 0.87 (9H, s, SiC(CH_3), 1.15 (3H, d, *J* = 6.1 Hz, 2-CH_3), 1.18 (3H, d, *J* = 6.9 Hz, 3-CH_3), 1.95 (3H, s, CO-CH_3), 2.65 (1H, dq, *J* = 6.8 Hz, 2-CH), 3.0 (2H, m, S-CH_2), 3.42 (2H, m, N-CH_2), 4.01 (1H, dq, *J* = 6.1 Hz, 3-CH), 5.86 (1H, br s, NH); δ_C (400 MHz, $CDCl_3$); -4.9 (Si-CH_3), -4.3 (Si-CH_3), 13.4 (2-CH_3), 18.0 (3-CH_3), 21.8 (CO-CH_3), 23.2 (Si-C), 25.7 (SiC(CH_3)$_3$), 28.3 (S-CH_2), 39.8 (NH-CH_2), 56.4 (2-CH), 69.9 (3-CH), 170.2 (6-CO), 202.8 (CO); vmax (neat) / cm-1: 3279 (NH), 2955 (CH), 2929 (CH), 2886 (CH), 2856 (CH), 1682 (CO), 1654 (CO), 1550 (NH), 1374 (CH), 1360 (CH); *m/z* (ESI) 356.17 [M+Na^+]; Found (ESI) 356.1671 ($C_9H_{17}NO_3S$ requires 356.1686).





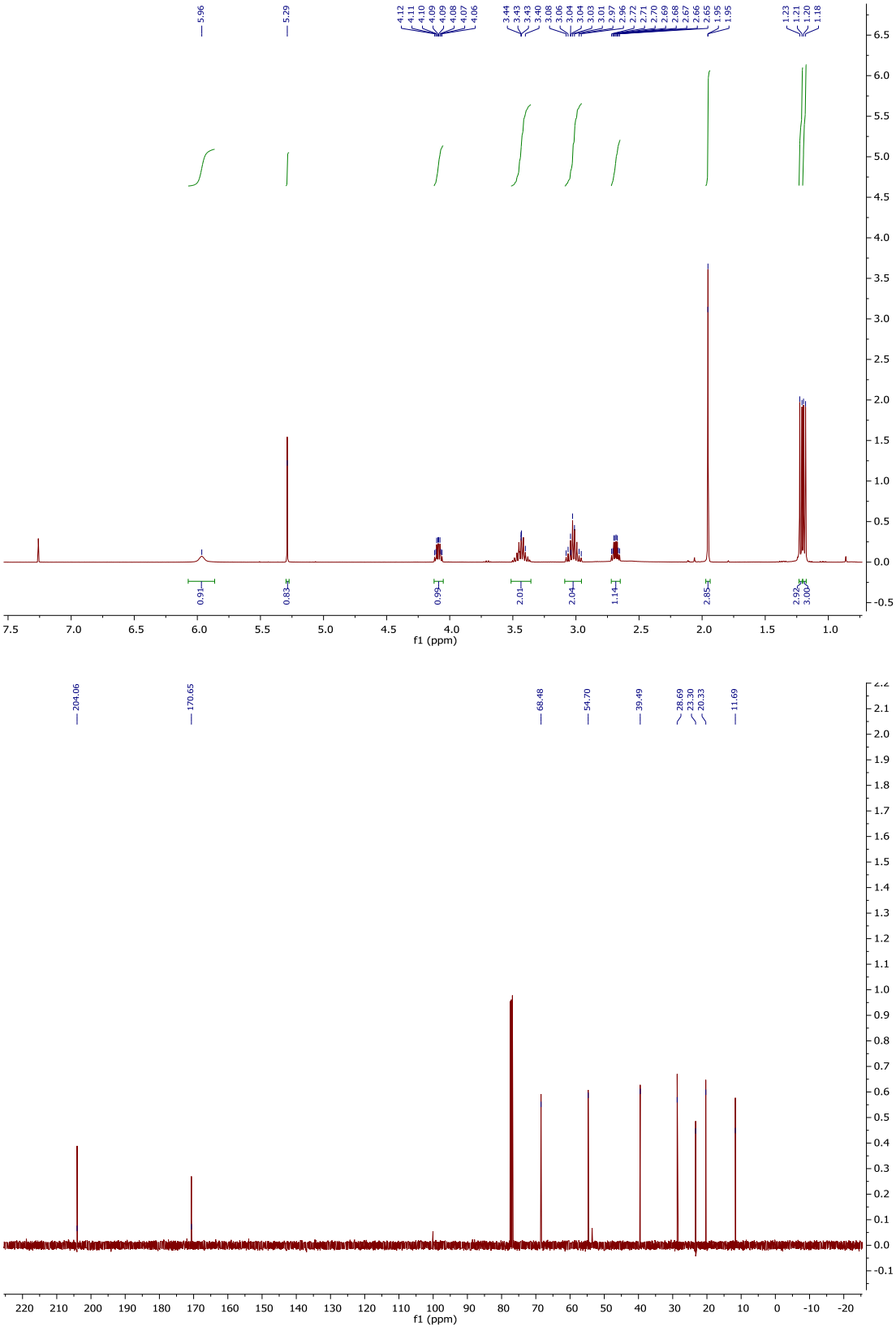
^{1}H 400 MH and ^{13}C NMR, 100 MHz, of (3*R*,2*S*) **162** in $CDCl_3$.

7.8.3.7 (2*S*,3*R*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 117



117

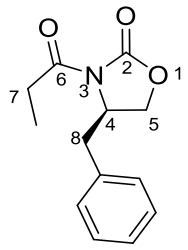
TBDMS-ether **162** (119 mg, 0.35 mmol) was dissolved in THF (0.9 mL), water (0.9 mL) and acetic acid (2.6 mL) and stirred at room temperature for 5 days. The reaction was diluted with water (25 mL) and the aqueous layer washed with CH_2Cl_2 (25 mL). The aqueous layer was concentrated to give the title compound **117** (37 mg, 0.16 mmol, 47%) as a colourless oil. $[\alpha]_D^{22.0}$+3.0 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$): 1.19 (3H, d, *J* = 6.4 Hz, 2-CH_3), 1.22 (3H, d, *J* = 7.1 Hz, 3-CH_3), 1.95 (3H, s, CO-CH_3), 2.69 (1H, qd, *J* = 7.0, 4.0 Hz, 2-CH), 3.04 (2H, m, S-CH_2), 3.43 (2H, m, NH-CH_2), 4.09 (1H, qd, *J* = 6.3, 4.1 Hz, 3-CH), 5.96 (1H, br s, NH); δ_C (400 MHz, CDCl3): 11.7 (2-CH_3), 20.3 (3-CH_3), 23.3 (CO-CH_3), 28.7 (S-CH_2), 39.5 (N-CH_2), 54.7 (2-CH), 68.5 (3-CH), 170.7 (N-CO), 204.1 (S-CO); ν_{max} (neat) / cm^{-1}: 3288 (NH), 3135 (OH), 2930 (CH), 1655 (NCO), 1548 (SCO); *m/z* (ESI) 242.08 [$M+Na^+$]; Found (ESI) 242.0816 ($C_9H_{17}NaNO_3S$ requires 242.0821).



^{1}H 400 MHz and ^{13}C NMR, 100 MHz, of (3*R*, 2*S*) **117** in $CDCl_3$.

7.8.4 (2*R*,3*S*)-3-Hydroxy-2-methylbutanoyl-*N*-acetylecysteamine

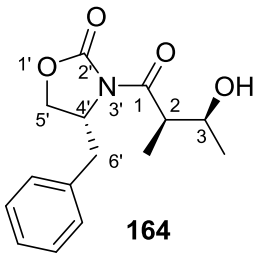
7.8.4.1 (*R*)-4-Benzyl-3-propionyloxazolidine-2-one 130[177]



130

(*R*)-4-benzyl-2-oxazolidinone (4.0 g, 22.6 mmol) was dissolved in dry THF (80 mL) under N_2 and cooled to -78 °C, n-BuLi (1.5 M in hexane) (14.9 mL, 22.6 mmol) was added dropwise and the solution stirred for 20 minutes. Propionyl chloride (2.20 mL, 24.8 mmol) was added and stirred for 3 hours before being warmed to room temperature. The reaction was quenched with NH_4Cl aq. (20 mL) and the majority of the THF was removed *in vacuo*. CH_2Cl_2 (40 mL) was added and the solution washed with 10% aq NaOH. (20 mL) The aqueous layer was extracted with CH_2Cl_2 (2 x 20 ml). The organic layers were combined, washed (brine), dried ($MgSO_4$), filtered and concentrated *in vacuo* to give the title compound **130** (5.22 g, 22.4 mmol, 99% yield) as a white solid. $[\alpha]_D^{24.0}$-62.8 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -65.1 (c = 1.01, $CHCl_3$)][178]; m.p. =41.7 - 42.8 °C; δ_H (400 MHz, $CDCl_3$): 1.20 (3H, t, *J* = 7.3 Hz, 7-CH_3), 2.77 (1H, dd, *J* = 9.7, 13.2 Hz, 8-CH_2), 2.96 (2H, dq, *J* = 7.7, 9.6 Hz, 7-CH_2), 3.30 (1H, dd, *J* = 3.3, 13.2 Hz, 8-CH_2), 4.17 (2H, m, 5-CH_2), 4.67 (1H, m, 4-CH), 7.19-7.34 (5H, m, Ar-H); ν_{max} (neat) / cm^{-1}: 2982 (CH), 2941 (CH), 1779 (C=O), 1698 (NC=O), 1349 (CN), 1213 (CO). ^1H NMR data in accordance with the literature.[177]

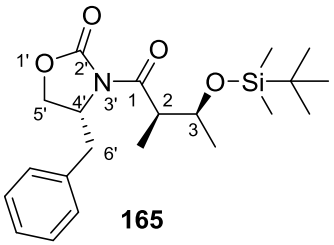
7.8.4.2 (4*R*)-3-[(2*R*,3*S*)-3-Hydroxy-2-methylbutyl]-4-phenylmethyl-2-oxazolidinone **164**[122]



164

(*R*)-4-Benzyl-3-propionyloxazolidin-2-one **130** (860 mg, 3.69 mmol) was dissolved in anhydrous CH_2Cl_2 (10 mL) and cooled to -78 °C under N_2. Dibutyl boron triflate (1 M in CH_2Cl_2, 7.4 mL, 7.58 mmol) was added dropwise over 1 hour. After 30 minutes triethylamine (1.28 mL, 9.23 mmol) was then added dropwise and the solution stirred for 1 hour at -78 °C, then 100 minutes at 0 °C. The solution was cooled to -78 °C and a solution of acetaldehyde (2.4 mL, 42.9 mmol) in dry CH_2Cl_2 (4 mL) also at -78 °C was added slowly. After stirring for 2 hours at -78 °C and then 100 minutes at 0 °C, 1 M phosphate buffer (pH 7, 28 mL) and MeOH (28 mL) were added, the resulting mixture was stirred vigorously in an ice bath as 30% H_2O_2 (14 mL) was added dropwise. After 30 minutes at 10-15 °C, the solvent was removed in vacuo and the resulting residue was partitioned between saturated $NaCO_3$ aq. (20 mL) and CH_2Cl_2 (20 mL). The aqueous layer was extracted with CH_2Cl_2 (2 x 8 mL) and the combined organic layers were dried ($MgSO_4$), filtered, and concentrated *in vacuo* to give a crude pale yellow oil. The crude material was purified by flash chromatography (SiO2, 20% EtOAc in petroleum ether, visualized with $KMnO_4$) to give the aldol adduct **164** as a colourless oil (693 mg, 2.10 mmol, 68% yield). $[\alpha]_D^{23.0}$ -47.2 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -54.8 (c = 0.96, $CHCl_3$)][122]; δ_H (400 MHz, $CDCl_3$): 1.21 (3H, d, *J* = 6.4 Hz, 2-CH_3), 1.27 (3H, d, *J* = 7.1 Hz, 3-CH_3), 2.79 (1H, dd, *J* = 13.4, 9.4 Hz, 6'-CH_2), 3.25 (1H, dd, *J* = 13.4, 3.4 Hz, 6'-CH_2), 3.74 (1H, qd, *J* = 7.1, 3.0 Hz, 2-CH), 4.21 (3H, m, 5'-CH & 3-CH), 4.69 (1H, m, 4'-CH), 7.20-7.35 (5H, m, Ar-H); ν_{max} (neat) / cm^{-1}:3509 (OH), 2977 (CH), 2926 (CH), 1772 (C=O), 1690 (NC=O), 1349 (C-N), 1208 (COC). ^{1}H NMR data in accordance with the literature.[122]

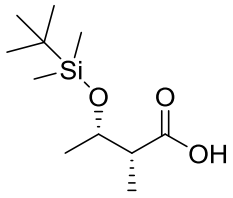
7.8.4.3 (4*R*)-3-[(2*R*,3*S*)-3- *tert*-Butyldimethylsilyloxy-2-methylbutyl]-4-phenylmethyl-2-oxazolidinone 165 [121]



165

Aldol adduct **164** (693 mg, 2.49 mmol) was dissolved in anhydrous CH_2Cl_2 (20 mL) under N_2. TBDMSCl (753 mg, 4.99 mmol), imidazole (678 mg, 9.96 mmol) and DMAP (3 crystals) were added and the reaction mixture stirred overnight. The reaction was quenched with $NaHCO_3$ (10 mL) and extracted with CH_2Cl_2 (3 x 25 mL). The organic extracts were dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 10% EtOAc in petroleum ether, visualised with $KMnO_4$) gave **165** (729 mg, 1.86 mmol, 75%) as a colourless crystalline solid. $[\alpha]_D^{22.0}$-43.1 (*c* 1.0, $CHCl_3$), Enantiomer [lit $\alpha]_D$ +49.2 (c = 1.0, $CHCl_3$)][121]; δ_H (400 MHz, $CDCl_3$); 0.02 (3H, s, Si-CH_3), 0.05 (3H, s, Si-CH_3), 0.88 (9H, s, SiC-CH_3), 1.19 (3H, d, *J* = 6.1 Hz, 2-CH_3), 1.22 (3H, d, *J* = 6.8 Hz, 3-CH_3), 2.77 (1H, dd, *J* = 13.4, 9.6 Hz, 6'-CH_2), 3.29 (1H, dd, *J* = 13.4, 3.24 Hz, 6'-CH_2), 3.83 (1H, dq, *J* = 6.7, 5.9 Hz, 2-CH), 4.09 (1H, p, *J* = 6.0 Hz, 3-CH), 4.17 (2H, m, 5'-CH_2), 4.62 (1H, m, 4'-CH), 7.20-7.36 (5H, m, Ar-H); δ_C (400 MHz, $CDCl_3$): -4.9 ($SiCH_3$), -4.4 ($SiCH_3$), 12.6 (2-CH_3), 21.7 (3-CH_3), 25.8 (SiC($\underline{CH_3}$)$_3$), 37.8 (SiC), 45.0 (Ar-CH_2), 55.7 (NCH), 66.0 (OCH_2), 69.7 (3-CH), 127.4 (Ar-C), 129.0 (Ar-C), 129.5 (Ar-C), 135.4 (O-CO), 175.4 (1-CO). ^1H and ^{13}C NMR data in accordance with the literature.[121]

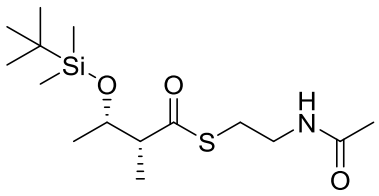
7.8.4.4 (2*R*,3*S*)-3-*tert*-Butyldimethylsilyloxy-2-methylbutanoic acid 166[174]



166

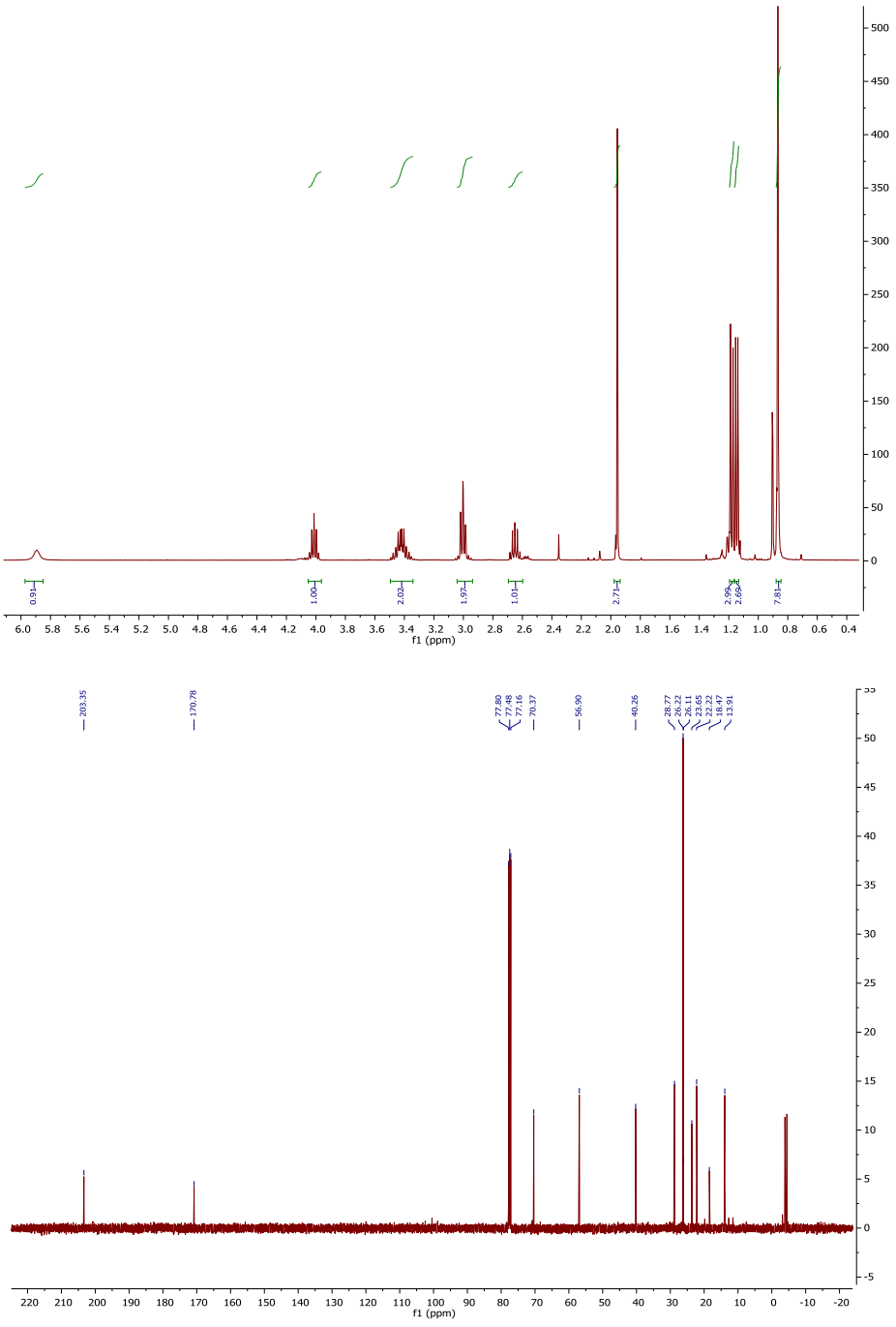
A solution of protected aldol **165** (718 mg, 1.83 mmol) dissolved in THF/H_2O (60 mL/9 mL) was cooled to 0 °C. Hydrogen peroxide (9.12 mmol, 1.04 mL, 30% w/v H_2O_2) and LiOH (153 mg, 3.66 mmol) were added. The reaction was warmed to room temperature. After 4 hours the excess peroxides were quenched using 1.5 M sodium sulfate (Na_2SO_3) aq. and the THF removed *in vacuo*. The resulting slurry was acidified to pH 2 using 2M HCl and extracted with CH_2Cl_2 (3 x 30 mL). The organic layers were combined, washed with brine (30 mL), dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification of the crude material by flash chromatography (SiO_2, 10 EtOAc/petroleum ether / 0.1% Acetic acid, visualised with $KMnO_4$) afforded the title compound **166** (230 mg, 0.99 mmol, 54%) as a colourless oil. $[\alpha]_D^{23.0}$ -35.3 (*c* 1.0, $CHCl_3$), [lit $\alpha]_D$ -22.3 (c = 1.07, $CHCl_3$)][174]; δ_H (400 MHz, $CDCl_3$); 0.12 (3H, s, Si-CH_3), 0.13 (3H, s, Si-CH_3), 0.91 (9H, s, SiC-$(CH_3)_3$), 1.14 (3H, d, *J* = 7.1 Hz, 2-CH_3), 1.19 (3H, d, *J* = 6.3 Hz, 3-CH_3), 2.59 (1H, dq, *J* = 7.1, 4.8 Hz, 2-CH), 4.07 (1H, dq, *J* = 6.3, 4.8 Hz, 3-CH); δ_C (400 MHz, $CDCl_3$);-5.0 (Si-CH_3), -4.4 (Si-CH_3), 12.1 (2-CH_3), 18.1 (Si-C), 20.2 (3-CH_3), 25.8 (SiC$(CH_3)_3$), 46.5 (2-CH), 70.1 (3-CH), 177.8 (CO); ν_{max} (neat) / cm^{-1}: 2956 (CH), 2887 (CH), 1705 (C=O), 831 ($SiCH_3$). ^{1}H NMR data in accordance with the literature.[174]

7.8.4.5 (2*R*,3*S*)-3-*tert*-Butyldimethylsilyloxy-2-methylbutanoyl-*N*-acetylcysteamine **167**



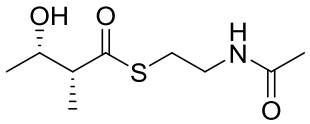
167

Acid **166** (185 mg, 0.80 mmol) was dissolved in anhydrous CH_2Cl_2 (12 mL) at 0 °C under nitrogen EDCI (eq, 173 mg, 1.11 mmol) and DMAP (1.0 eq, 48 mg, 0.04 mmol) was added. After stirring at 0 °C for 15 minutes, HSNAC **151** (133 mg, 1.12 mmol) in anhydrous CH_2Cl_2 (10 mL) was added. The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (15 mL) and the layers separated. The aqueous phase was extracted with CH_2Cl_2 (3 x 25 mL). The organics were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography gave the title compound **167** (193 mg, 0.58 mmol, 72%) as a colourless oil. $[\alpha]_D^{21.0}$ -11.0 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 0.04 (3H, s, $SiCH_3$), 0.06 (3H, s, $SiCH_3$), 0.88 (9H, s, $SiC\text{-}(CH_3)_3$), 1.15 (3H, d, *J* = 6.2 Hz, 2-CH_3), 1.19 (3H, d, *J* = 6.8 Hz, 3-CH_3), 2.65 (1H, m, 2-CH), 3.01 (2H, td, *J* = 6.4, 1.2 Hz, S-CH_2), 3.43 (2H, m, N-CH_2), 4.02 (1H, p, *J* = 6.1 Hz, 3-CH), 5.80 (1H,brs,NH); δ_C (400 MHz, $CDCl_3$); -5.1 (Si-CH_3),), -4.2 (Si-CH_3), 13.6 (2-CH_3), 18.1 (Si-$\underline{C}(CH_3)_3$), 20.9 (3-CH_3), 23.4 (CO-$\underline{C}H_3$), 25.9 ($SiC(CH_3)_3$), 28.5 (S-CH_2), 39.8 (N-CH_2), 56.9 (2-CH), 70.3 (3-CH), 170.4 (N-CO), 203.2 (S-CO); ν_{max} (neat) / cm-1: 3286 (NH), 2955 (CH), 2930 (CH), 2857 (CH), 1683 (C=O), 1656 (C=O), 1551 (NH), 836 ($SiCH_3$); *m/z* (ESI) 356.17 [$M+Na^+$]; Found (ESI) 356.1689 ($C_9H_{17}NO_3S$ requires 356.1686).



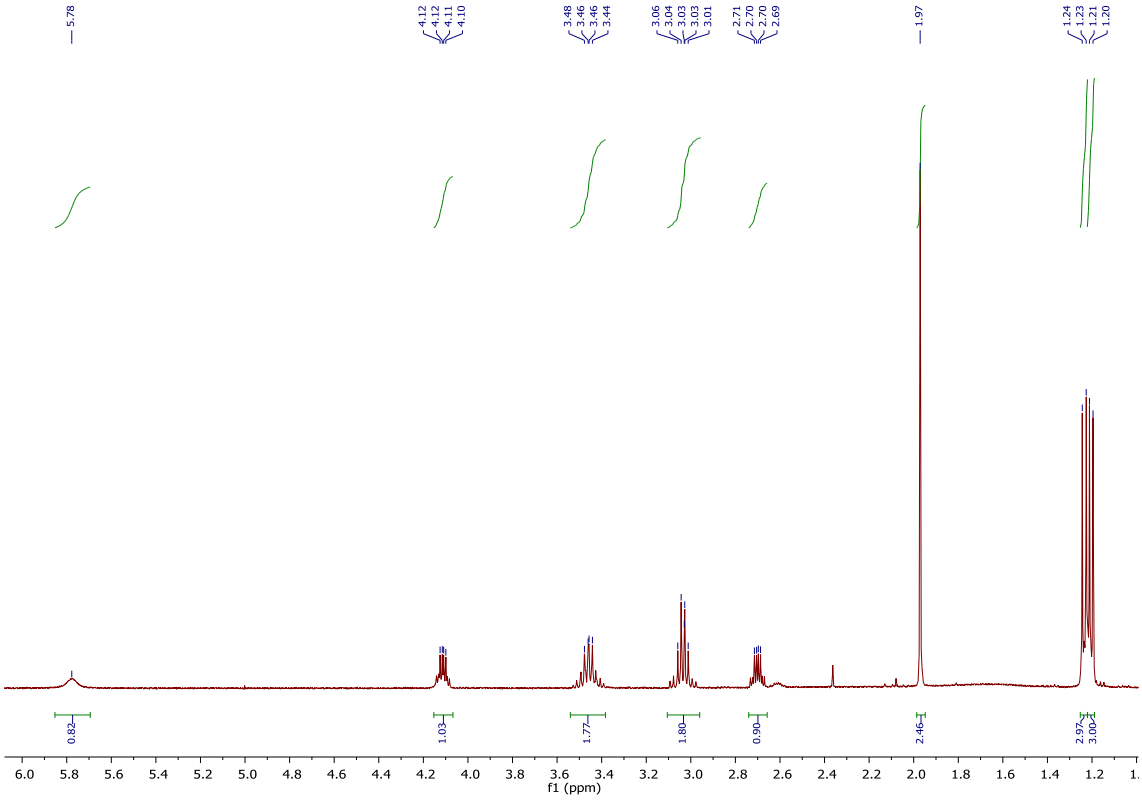
^{1}H, 400 MHz and ^{13}C NMR, 100 MHz, of (3*S*,2*R*) **167** in $CDCl_3$.

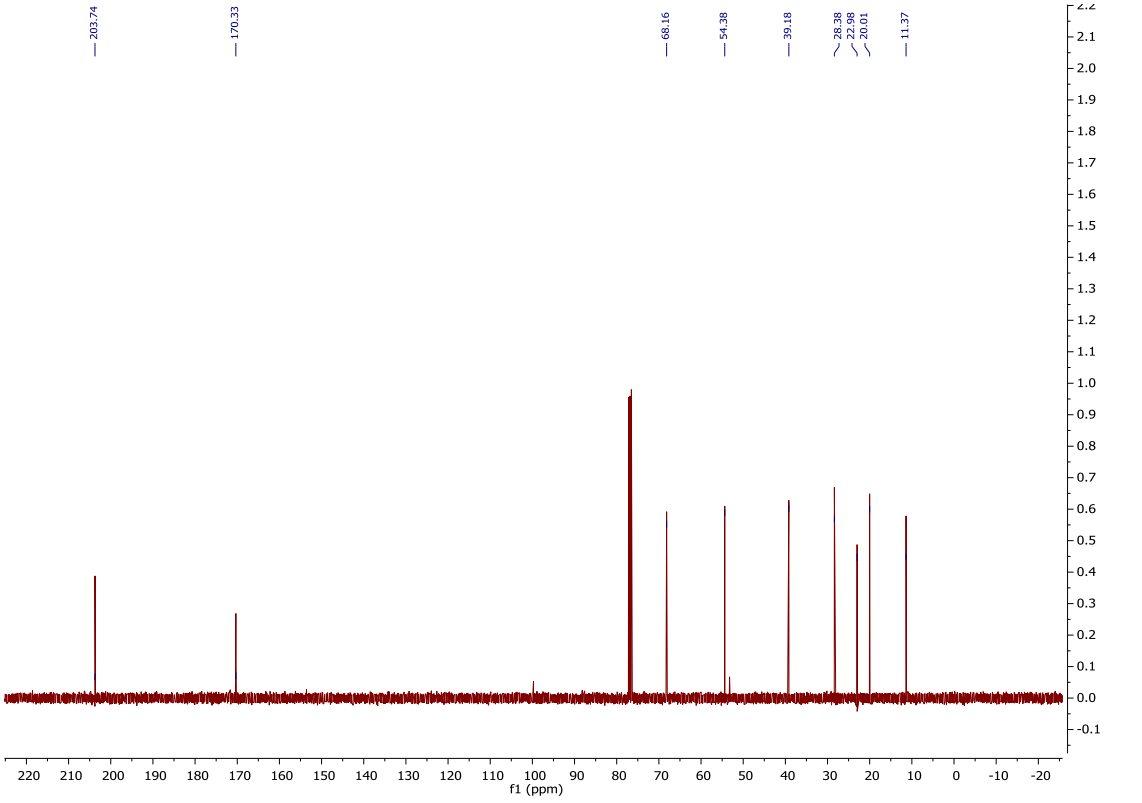
7.8.4.6 (2*R*,3*S*)- 3-Hydroxy-2-methylbutanoyl-*N*-acetylcysteamine 116



116

TBDMS-ether **167** (180 mg, 0.54 mmol) was dissolved in THF (1.5 mL), water (1.5 mL) and acetic acid (5 mL) and stirred at room temperature for 3 days. The reaction was diluted with water (15 mL) and the aqueous layer washed with CH_2Cl_2 (15 mL). The aqueous layer was concentrated to give the title compound **116** (62.0 mg, 0.28 mmol, 52%) as a colourless oil. $[\alpha]_D^{22.0}$ -4.9 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 1.20 (3H, d, *J* = 6.4 Hz, 2-CH_3), 1.23 (3H, d, *J* = 7.1 Hz, 3-CH_3), 1.97 (3H, s, CO-CH_3), 2.70 (1H, qd, *J* = 7.1, 4.0 Hz, 2-CH), 3.04 (2H, m, S-CH_2), 3.46 (2H, m, N-CH_2), 4.11 (1H, qd, *J* = 6.4, 4.0 Hz, 3-CH), 5.78 (1H, br s, NH); δ_C (400 MHz, CDCl3); 11.4 (2-CH_3), 20.1 (3-CH_3), 23.0 (CO-CH_3), 28.4 (S-CH_2), 39.2 (N-CH_2), 54.4 (2-CH), 68.2 (3-CH), 170.3 (N-CO), 203.7 (S-CO); ν_{max} (neat) / cm^{-1}:3299 (NH), 3134 (OH), 2974 (CH), 2934 (CH), 1653 (C=O), 1549 (C=O), 1451 (NH); *m/z* (ESI) 242.08 [$M+Na^+$]; Found (ESI) 242.0813 ($C_9H_{17}NaNO_3S$ requires 242.0821).

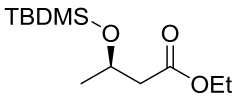




^{1}H 400 MHz and ^{13}C NMR, 100 MHz, of (3*S*, 2*R*) **116** in $CDCl_3$.

7.8.5 (3*R*)-Hydroxybutanoyl-*N*-acetylcysteamine

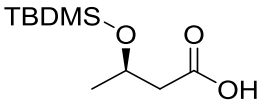
7.8.5.1 Ethyl (3*R*)-(*tert*-Butyldimethylsilyloxy)-butanoate 157[179]



157

Pyridine (1.20 mL, 16.0 mmol) was added to ethyl-(3*R*)-3-hydroxybutyrate **141** (1.0 g, 7.56 mmol) dissolved in CH_2Cl_2 (20 mL) and cooled to 0 °C. *tert*-Butyldimethylsilyl triflate (2.4 mL, 10.6 mmol) was added dropwise. The mixture was allowed to warm to room temperature, stirred overnight and quenched with sat. $NaHCO_3$ aq. (20 mL). The layers were separated and the aqueous layer extracted with CH_2Cl_2 (3 x 20 mL), dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 5% EtOAc/ petroleum ether visualised with $KMnO_4$) gave the title compound **157** (1.15 g, 4.96 mmol, 67%) as a colourless oil. $[\alpha]_D^{27.0}$ -23.0 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -27.5 (c = 1.02, $CHCl_3$)].[180]; δ_H (400 MHz, $CDCl_3$); 0.03 (3H, s, Si-CH_3), 0.05 (3H, s, Si-CH_3), 0.85 (9H, s, Si-C(CH_3)$_3$), 1.18 (3H, d, *J* = 6.1 Hz, 3-CH_3), 1.25 (3H, t, *J* = 7.1 Hz, $CH_2\underline{CH_3}$), 2.35 (1H, dd, *J* = 5.4, 14.3 Hz, 2-CH_2), 2.46 (1H, dd, *J* = 7.7, 14.3 Hz, 2-CH_2), 1.44 (2H, dq, *J* = 4.2, 3.9 Hz, O-CH_2), 4.27 (1H, m, 3-CH); ν_{max} (neat) / cm^{-1}: 2957 (CH), 2930 (CH), 2896 (CH), 2857 (CH), 1736 (CO); ^{1}H NMR data in accordance with the literature.[179]

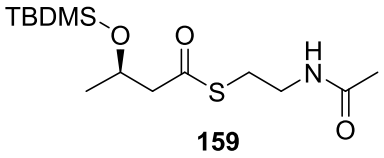
7.8.5.2 (3*R*)-3-(*tert*-Butyldimethylsilyloxy)-butanoic acid 158[181]



158

Ester **157** (517 mg, 3.78 mmol) was dissolved in THF (2.5 mL), MeOH (4 mL) and lithium hydroxide (271 mg, 11.3 mmol) in water (0.75 mL) were added. The reaction mixture was heated to 60 °C overnight. The reaction was diluted with water (5 mL) and extracted with Et_2O (2 x 5 mL). The aqueous layer was acidified to pH 3 with H_2SO_4 (6 M) and further extracted with EtOAc (3 x 7 mL), the pH was adjusted to pH 3 as required. The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo* to give a pale yellow oil. Purification by flash chromatography (SiO_2, 10% EtOAc/ Petroleum ether/ 0.5% acetic acid visualised with $KMnO_4$) afforded the title compound **158** (299 mg, 1.37 mmol, 36%) as a colourless oil. $[\alpha]_D^{22.8}$ -12.0 (c 1.0, $CHCl_3$), [lit $[\alpha]_D$ -10.8 (c 0.25, $CHCl_3$)][181]; δ_H (400 MHz, $CDCl_3$); 0.06 (3H, s, Si-CH_3), 0.08 (3H, s, Si-CH_3), 0.87 (9H, s, Si-C(CH_3)$_3$), 2.47 (1H, d, *J* = 2.8 Hz, 2-CH_2), 2.49 (1H, d, *J* = 2.7 Hz, 2-CH_3), 4.28 (1H, sext., *J* = 6.1 Hz, 3-CH); ν_{max} (neat) / cm^{-1}: 2955 (CH), 2929 (CH), 2896 (CH), 2857 (CH), 1710 (CO) 1472 (CH). ^{1}H NMR data in accordance with the literature.[181]

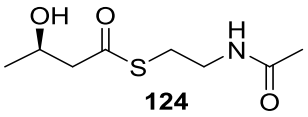
7.8.5.3 (3*R*)-3-(*tert*-Butyldimethylsilyloxy)-butanoyl-*N*-acetylcysteamine 159



159

Acid **158** (200 mg, 0.98 mmol) was dissolved in anhydrous CH_2Cl_2 (12 mL) at 0 °C under nitrogen. EDCI (213 mg, 1.37 mmol) and DMAP (16 mg, 0.14 mmol) were added. After stirring at 0 °C for 15 minutes, HSNAC **151** (163 mg, 1.37 mmol) was added in anhydrous CH_2Cl_2 (5 mL). The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (5 mL) and the layers separated. The aqueous phase was extracted with CH_2Cl_2 (3 x 15 mL). The organics were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 3% MeOH/ CH_2Cl_2 visualised with $KMnO_4$) gave the title compound **159** (174 mg, 0.54 mmol, 56%) as a colourless oil. $[\alpha]_D^{23.0}$ -26.8 (c 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 0.02 (3H, s, Si-CH_3), 0.05 (3H, s, Si-CH_3), 0.85 (9H, s, Si-C(CH_3)$_3$), 1.18 (3H, d, *J* = 6.1 Hz, 3-CH_3), 1.96 (3H, s, CO-CH_3), 2.59 (1H, dd, *J* = 5.1, 14.7 Hz, 2-CH_2), 2.75 (1H, dd, *J* = 7.8, 14.4 Hz, 2-CH_2), 3.01 (2H, m, S-CH_2), 3.43 (2H, m, N-CH_2), 4.29 (1H, m, 3-CH), 5.91 (1H, br s, NH); δ_C (400 MHz, CDCl3): -4.9 ($SiCH_3$), -4.8 ($SiCH_3$), 18.1 (3-CH_3), 23.6 (SiC($\underline{C}H_3$), 23.9 ($COCH_3$) 25.9 (SiC(CH_3), 28.9 (NCH_3), 39.8 (SCH_2), 53.9 (2-CH_2), 66.1 (3-CH), 170.5 (NCO), 196.5(SCO); ν_{max} (neat) / cm^{-1}: 3288 (NH), 2965 (CH), 2929 (CH), 2895 (CH), 2856 (CH), 1688 (CO), 1656;); *m/z* (ESI) 342.15 [M+Na^+]; Found (ESI) 342.1535 ($C_{14}H_{29}NO_3SSi$ requires 342.1530).

7.8.5.4 (3*R*)-3-Hydroxy-butanoyl-*N*-acetylcysteamine 124

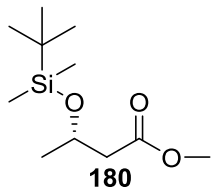


124

Butanoate **159** (100 mg, 0.30 mmol) was dissolved in THF (0.9 mL), water (0.9 mL) and acetic acid (2.6 mL) and stirred at room temperature for 5 days. The reaction was diluted with water (25 mL) and the aqueous layer washed with CH_2Cl_2 (25 mL). The aqueous layer was concentrated to give the title compound **124** (28.0 mg, 0.17 mmol, 58%) as a colourless oil. $[\alpha]_D^{23.7}$ -23.2 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 1.23 (3H, d, *J* = 6.3 Hz, 3-CH_3), 1.96 (3H, s, $COCH_3$), 2.71 (1H, d, *J* = 3.8 Hz, 2-CH_2), 2.72 (1H, d, *J* = 0.8 Hz, 2-CH_2), 3.03 (2H, dq, *J* = 3.7, 6.4 Hz, S-CH_2), 3.43 (2H, q, *J* = 6.3 Hz, N-CH_2), 4.24 (1H. m, 3-CH), 6.03 (1H, br s, NH); δ_C (400 MHz, $CDCl_3$); 22.9 (3-CH_3), 23.7 (CO-CH_3), 28.9 (S-CH_2), 39.4 (N-CH_2), 52.6 (2-CH_2), 170.8 (NCO). 199.4 (SCO); ν_{max} (neat) / cm^{-1}: 3288 (NH), 3074 (OH), 2967 (CH), 2920 (CH), 1654 (C=O), 1550 NH ;); *m/z* (ESI) 228.07 [M+Na^+]; Found (ESI) 228.0668 ($C_8H_{15}NO_3S$ requires 228.0665).

7.8.6 (3*S*)-Hydroxybutanoyl-*N*-acetylcysteamine

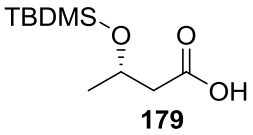
7.8.6.1 Methyl (3*S*)-3-(*tert*-Butyldimethylsilyloxy)butanoate 180[182]



180

Pyridine (4.10 mL, 17.8 mmol) was added to ethyl-(3*S*)-3-hydroxybutyrate (1.00 g, 8.47 mmol), dissolved in CH_2Cl_2 (20 mL) and cooled to 0 °C. *tert*-Butyldimethylsilyl triflate (2.4 mL, 10.6 mmol) was added dropwise. The mixture was allowed to warm to room temperature, stirred for 1 hour and quenched with sat. $NaHCO_3$ (20 mL). The layers were separated and the aqueous layer extracted with CH_2Cl_2 (3 x 20 mL), dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 5% EtOAc/ petroleum ether visualised with $KMnO_4$) gave the title compound **180** (1.38 g, 5.95 mmol, 70%) as a pale yellow oil. $[\alpha]_D^{24.4}$ +29.6 (*c* 1.0, $CHCl_3$), [lit $[\alpha]_D$ +30.3 (*c* 1.0, CH,Cl_3)][183]; δ_H (400 MHz, $CDCl_3$); 0.02 (3H, s, Si-CH_3), 0.04 (3H, s, Si-CH_3), 0.84 (9H, s, Si-C(CH_3)$_3$), 1.17 (3H, d, *J* = 6.1 Hz, 3-CH_3), 2.35 (^{1}H, dd, *J* = 5.8, 14.5 Hz, 2-CH_2), 2.45 (1H, dd, *J* = 7.7, 14.5 Hz, 2-CH_3), 3.64 (3H, s, O-CH_3), 4.26 (1H, m, 3-CH); ν_{max} (neat) / cm^{-1}: 2956 (CH), 2930 (CH), 2858 (CH). 1710 (C=O), 1001 (CO), 827 ($SiCH_3$). 1H NMR data in accordance with the literature.[182]

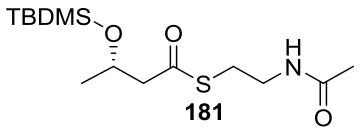
7.8.6.2 (3*S*)-3-(*tert*-Butyldimethylsilyloxy)butanoic acid 179[184]



179

Ester **180** (1.0 g, 4.31 mmol) was dissolved in THF (5 mL), MeOH (8 mL) and lithium hydroxide (542 mg, 12.9 mmol) in water (1.5 mL) was added. The reaction mixture was heated to 60 °C overnight. The reaction was diluted with water (7 mL) and extracted with Et_2O (2 x 10 mL). The aqueous layer was acidified to pH 3 with H_2SO_4 (3 M) and further extracted with EtOAc (3 x 15 mL), the pH was adjusted to pH 3 as required. The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo* to give a pale yellow oil. Purification by flash chromatography (SiO_2, 10% EtOAc/ petroleum ether/ 0.5% acetic acid visualised with $KMnO_4$) afforded the title compound **179** (299 mg, 1.37 mmol, 36%) as a colourless oil. $[\alpha]_D^{23.7}$ +11.7 (*c* 1.0, $CHCl_3$), [lit $[\alpha]_D$ +11.9 (*c* 1.29 , CH,Cl_3)][185]; δ_H (400 MHz, $CDCl_3$); 0.06 (3H, s, Si-CH_3), 0.08 (3H, s, Si-CH_3), 0.87 (9H, s, Si-C(CH_3)$_3$), 1.23 (3H, d, *J* = 6.2 Hz, 3-CH_3), 2.46 (1H, dd, *J* = 5.5, 13.6 Hz, 2-CH_2), 2.51 (1H, dd, *J* = 6.7, 13.6 Hz, 2-CH_2), 4.27 (1H, septet, *J* = 6.1 Hz, 3-CH); ν_{max} (neat) / cm^{-1}: 3358 (OH), 2925 (CH), 1689 (C=O). ^{1}H NMR data in accordance with the literature.[184]

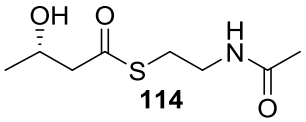
7.8.6.3(3*S*)-3-(*tert*-Butyldimethylsilyloxy)butanoyl-*N*-acetylcysteamine 181



181

Acid **179** (300 mg, 1.37 mmol) was dissolved in dry CH_2Cl_2 (19 mL) at 0 °C under nitrogen. EDCI (376 mg, 1.92 mmol) and DMAP (17.0 mg, 0.14 mmol) were added. After stirring at 0 °C for 15 minutes, HSNAC **151** (228 mg, 1.92 mmol) was added in anhydrous CH_2Cl_2 (7.5 mL). The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (15 mL) and the layers separated. The aqueous phase was extracted with CH_2Cl_2 (3 x 25 mL). The organics were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. Purification by flash chromatography (SiO_2, 3% MeOH/ CH_2Cl_2 visualised with $KMnO_4$) gave the title compound **181** (174 mg, 0.54 mmol, 56%) as a colourless oil. $[\alpha]_D^{24.4}$ +101.5 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 0.01 (3H, s, $SiCH_3$), 0.04 (3H, s, $SiCH_3$), 0.84 (9H, s, $SiC(CH_3)_3$), 1.17 (d, *J* = 6.1 Hz, 3-CH_3), 1.95 (3H, s, $COCH_3$), 2.58 (1H, dd, *J* = 5.0, 14.4 Hz, 2-CH_2), 2.74 (1H, dd, *J* = 7.4, 14.4 Hz, 2-CH_2), 3.01 (2H, qt, *J* = 6.4, 13.9 Hz, S-CH_2), 3.41 (2H, m, N-CH_2), 4.28 (1H, m, 3-CH), 5.84 (1H, br s, NH);); δ_C (400 MHz, CDCl3): -4.8 ($SiCH_3$), -4.9 ($SiCH_3$), 18.1 (3-CH_3), 23.4 (SiC($\underline{C}H_3$), 23.8 ($COCH_3$) 25.9 (Si$\underline{C}(CH_3)$, 28.8 (NCH_3), 39.8 (SCH_2), 53.9 (2-CH_2), 66.1 (3-CH), 170.5 (NCO), 196.1(SCO); ν_{max} (neat) / cm^{-1}: 3281 (NH), 2955 (CH), 2929 (CH), 2887 (CH), 2856 (CH), 1689 (CO), 1652 (CO), 1550 (NH), 1375 (CH), 1361 (CH); *m/z* (ESI) 342.15 [M+Na^+]; Found (ESI) 342.1545 ($C_{14}H_{29}NO_3SSi$ requires 342.1530).

7.8.6.4 (3*S*)-3-Hydroxy-butanoyl-*N*-acetylcysteamine 114



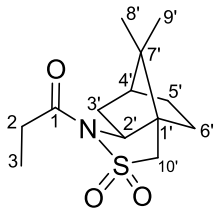
114

Butanoate **181** (119 mg, 0.54 mmol) was dissolved in THF (0.9 mL), water (0.9 mL) and acetic acid (2.6 mL) and stirred at room temperature for 3 days. The reaction was diluted with water (15 mL) and the aqueous layer washed with CH_2Cl_2 (15 mL). The aqueous layer was concentrated to give the title compound **114** (70 mg, 0.34 mmol, 63%) as a colourless oil. $[\alpha]_D^{24.6}$ +27.9 (*c* 1.0, $CHCl_3$); δ_H (400 MHz, $CDCl_3$); 1.24 (3H, d, *J* = 6.3 Hz, 3-CH_3), 1.97 (3H, s, $COCH_3$), 2.73 (2H, m, 2-CH_2), 3.04 (2H, td, *J* = 3.5, 6.3 Hz, S-CH_2), 3.45 (2H, qd, *J* = 2.9, 6.3 Hz, N-CH_2), 4.26 (1H, m, 3-CH), 5.81 (1H, br s, NH); δ_C (400 MHz, $CDCl_3$); 22.8 (CO=CH_3), 23.2 (3-CH_3), 28.9 (S-CH_2), 39.3 (N-CH_2), 52.6 (2-CH_2), 65.1 (3-CH), 170.7 (NHC=O), 199.4 (SC=O); ν_{max} (neat) / cm^{-1}: 3290 (OH), 3086 (NH), 2931 (CH), 1652 (CO), 1548 (NH), 1431 (CH), 1074 (CN);); *m/z* (ESI) 228.07 [M+Na^+]; Found (ESI) 228.0669 ($C_8H_{15}NO_3S$ requires 226.0665).

7.9 Synthesis of triketide substrate mimic

7.9.1 (3*R*,2*R*)-4-Methyl-3-hydroxy-2-methylhexanoyl-*N*-acetylcysteamine

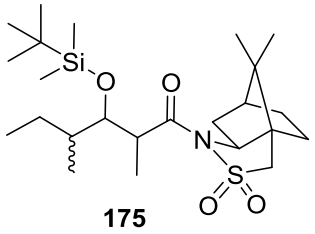
7.9.1.1 (1*R*,2*S*)-*N*-Propionyl-bornane-10-2-sultam 169[186]



169

(1*R*)-(+)-camphorsultam **172** (1 e.q., 5.0 g, 23 mmol) in dry toluene (40 mL) was added dropwise to a suspension of sodium hydroxide (60% dispersion in oil, 1.8 e.q., 1.4g, 41.8 mmol) in dry toluene (15 mL) under N_2. After stirring for 1 hour, propionyl chloride (1.1 e.q., 2.2 mL, 25.3 mmol) was added dropwise and stirred at room temperature for 2 hours. The reaction was quenched with saturated aq. NH_4Cl (150 mL) and stirred for 35 minutes. The aqueous layer was extracted with ethyl acetate (3 x 50 mL). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo* to give a white solid. The crude solid was recrystallized in methanol and the isolated solid was redissolved in CH_2Cl_2 to remove any residual methanol. (1*R*,2*S*)-*N*-Propionyl-bornane-10-2-sultam **169** (5.63 g, 20.7 mmol, 90%) was isolated as a colourless solid. mp. 142.3-143.5 °C; $[\alpha]_D^{23.7}$ +114.2 (*c* 1.0, $CHCl_3$), [lit $[\alpha]_D$ +116.5.0 (*c* 1.00 , CH,Cl_3)][186]; δ_H 400 MHz, $CDCl_3$); 0.97 (3H, s, 8'-CH_3), 1.18 (6H, m, 9'-CH_3 & 3-CH_3), 1.38 (2H, m, 5'-CH_2 & 6-CH_2), 1.90 (3H, m, 4'-CH, 5'-CH_2, 6'-CH_2), 2.09 (2H, m, 3'-CH_2), 2.75 (2H, qd, *J* = 7.5, 5.7 Hz, 2-CH_2), 3.46 (2H, q, *J* = 13.5 Hz, 10'-CH_2), 3.86 (1H, dd, *J* = 7.5, 5.0 Hz, 2'-CH); δc (400 MHz, $CDCl_3$); 8.5 (3-CH_3), 20.0 (8'-CH_3), 20.9 (9'-CH_3), 26.6 (2-CH_2), 29.1 (5'-CH_2), 33.0 (6'-CH_2), 38.6 (3'CH_3), 44.8 (4'-CH), 47.9 (7'-C), 48.6 (1'-CH), 53.7 (10'-CH_2), 65.4 (2'-CH); ν_{max} (neat) / cm^{-1}: 2953 (CH), 2922 (CH), 2881 (CH), 1687 (CO), 1327 (CN), 770 (C-S), 559 (CO). ^{1}H NMR data in accordance with the literature.[186]

7.9.1.2 (1*R*,2*S*)-*N*-[(3'*R*,2'*R*)-4-Methyl-3-(*tert*-butyldimethylsilanyloxy)-2-methylhexanoyl]-borane-10-2-sultam 175[113]



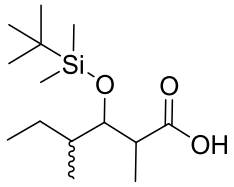
175

Triethylamine (freshly distilled over calcium hydride, 3 e.q., 11.524 mmol, 1.61 mL) was added to a solution of Propionyl sultam **169** (1 e.q., 3.84 mmol, 1.04 mg) in dry CH_2Cl_2 (8 mL) under N_2 at room temperature. After 5 minutes TBDMSOTf was added and the reaction stirred for 20 hours at room temperature. $NaHCO_3$ (10 mL) was added and the aqueous layer extracted with ethyl acetate (4 x 15 mL). The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo*. The residue was dissolved in dry CH_2Cl_2 (8 mL) under N_2 and cooled to -78 °C. Methylbutaldehyde (2.2 e.q., 8.45 mmol, 0.80 mL) was added slowly. After 10 minutes titanium (IV) chloride (1.0 M in CH_2CL_2, 2.2 e.q., 8.45 mml, 8.45 mL) was added dropwise. The reaction was stirred for 2 hours at -78 °C. The reaction was quenched with saturated NH_4Cl (15 mL) and extracted with CH_2CL_2 (4 x 20 mL). Organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo.* Purification by flash chromatography (SiO_2, 5-15% EtOAc/ Petrol visualised with $KMnO_4$) afforded the partially pure compound **174** (0.99g, 2.77 mmol, 72%) as a white solid. δ_H (400 MHz, $CDCl_3$); 0.89 (6H, m, 4-CH_3 & 6-CH_3), 0.97 (3H, s, 8'CH_3), 1.15 (3H, d, *J* = 6.6 Hz, 2-CH_3), 1.19 (3H, s, 9'-CH_3), 1.24-1.28 (4H, m, 5-CH_2, 6'-CH_2, 5'-CH_2), 1.85-1.91 (4H, m, 4-CH, 6'CH_2, 5'-CH_2, 4'-CH_2), 2.05-2.19 (2H, m, 3-CH_3), 3.28 (1H, p, *J* = 6.6 Hz, 2-CH), 3.50 (2H, q, *J* = 13.8 Hz, 10'CH_2), 3.61 (1H, m, 3-CH), 3.90 (1H, dd, *J* = 7.9, 4.7 Hz, 2'-CH). enantiomer ^{1}H NMR data in accordance with the literature.[113]

Aldol **174** (1 e.q., 1.0 g, 2.79 mmol) was dissolved in CH_2Cl_2 (11 mL) and cooled to 0 °C. Pyridine (2 e.q., 0.5 mL, 6.14 mmol) was added, followed by TBDMSOTf (1.5 e.q., 0.86 mL, 3.75 mmol). After stirring at room temperature for 2 hours, the reaction was quenched with sat. ammonium chloride (15 mL). The aqueous layer was extracted with CH_2Cl_2 (3 x 50 mL), organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo* to give a colourless solid. Purification by flash chromatography (SiO_2, 10-15% EtOAc/ Petrol visualised with $KMnO_4$) afforded the title compound **175** (761 mg, 1.62 mmol, 58%) as a colourless solid. mp. 104.7- 105.8 δ_H (400 MHz, $CDCl_3$); 0.05 (3H, s, Si-CH_3), 0.08 (3H, s, Si-CH_3), 0.81 (3H, t, *J* = 7.3 Hz, 6-CH_3), 0.87 (3H, s, 8'-CH_3), 0.88 (SiC(CH_3), 0.95 (3H, 9'-CH_3), 1.15 (3H, d, *J* = 4.0 Hz, 4-CH_3), 1.18 (3H, d, *J* = 6.6 Hz, 2-CH_3), 1.30-1.43 (4H, m, 5-CH_2, 5'-CH_2, 6'=CH_2), 1.87 (3H, m, 5'-CH_2, 6'-CH_2,4'-CH), 2.04 (3H, d, *J* = 7.4 Hz, 3'-CH_2), 3.29-3.49 (3H, m, 10'-CH_2 & 2-CH), 3.88 (1H, m, 3-CH), 4.06 (1H, dd, *J* = 4.8 Hz, 1.7, 2'-CH); δc (400 MHz, $CDCl_3$); -4.8

($SiCH_3$), 3.9 ($SiCH_3$), 12.3 (6-CH_3), 14.4 (4-CH_3), 18.4 (2-CH_3), 19.9 (9'-CH_3) 20.1 (8'CH_3), 26.0 (SiC($\underline{C}H_3$)$_3$, 26.1 (4-CH), 27.9 (5-CH_2), 33.0 (SiC), 38.6 (6'-CH_2), 44.8 (3'CH_2), 46.5 (2-CH), 47.7 (7'-C), 48.1 (4'-CH), 53.2 (10'CH_2), 65.7 (2'CH), 75.3 (3-CH), 174.2 (1-CO) ;ν_{max} (neat) / cm^{-1}; 2959.4 (CH), 2932 (CH), 2881 (CH), 1683 (CO), 1335 (SO), 1129 (SO), 327 (C-Si), 777 (C-S); *m/z* (ESI): 472.27 [M+Na^+]; Found (ESI) 494.2711 ($C_{24}H_{45}NNaO_5SSi$ requires 494.2731).

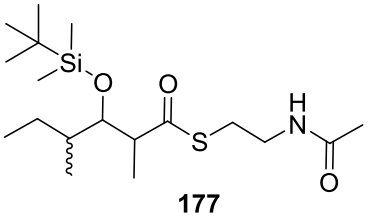
7.9.1.3 (3'*R*,2'*R*)-4-Methyl-3-(*tert*-butyldimethylsilanyloxy)-2-methylhexanoic acid 176



176

Aldol **175** (1 e.q., 700 mg, 1.48 mmol) was dissolved in THF (11.5 mL) and H_2O (4.6.mL). LiOH (2 e.q., 71 mg, 2.96 mmol) was added, followed by H_2O_2 (30% a.q., 7.4 e.q., 1.25 mL, 10.95 mmol). After 7 hours, the reaction was diluted with H_2O (15 mL) and washed with ether (8 mL). The aqueous layer was cooled to 0 °C and acidified to pH 5 with HCl (2 M) and extracted with ethyl acetate (6 x 15 mL). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo* to give a colourless oil. Purification by flash chromatography (SiO_2, 10% EtOAc/PE/ 0.1% Acetic acid, visualised with $KMnO_4$) gave the title compound **176** (226 mg, 0.82 mmol, 56%) as a colourless oil. δ_H (400 MHz, $CDCl_3$); 0.11 (3H, s, Si-CH_3), 0.14 (3, s, Si-CH_3) 0.92 (15H, m, SiC(CH_3)$_3$, 4-CH_3, 6-CH_3), 1.10 (1H, m, 4-CH); 1.23 (3H, d, *J* = 7.3 Hz, 2-CH_3), 1.55 (2H, m, 5-CH_2), 2.66 (1H, m, 2-CH), 3.76 (1H, 3-CH); δc (400 MHz, $CDCl_3$): -4.2 ($SiCH_3$), -3.7 ($SiCH_3$), 12.5 (6-CH_3), 14.3 (2-CH_3), 16.7 (4-CH_3), 19.7 (5-CH_2), 25.1 (SiC), 26.3 (SiC($\underline{C}H_3$)$_3$, 40.1 (3-CH), 43.0 (2-CH), 79.1 (3-CH), 173.5 (CO); ν_{max} (neat) / cm^{-1}; 3235 (OH), 2959 (CH). 2930 (CH), 1706 (CO), 833 (SiC); *m/z* (ESI): 297.18 [M+Na^+]; Found (ESI) 297.1853 ($C_{14}H_{30}NaO_3S$ requires 297.1865..

7.9.1.4 (3'*R*,2'*R*)-4-Methyl-3-(*tert*-butyldimethylsilanyloxy)-2-methylhexanoyl-*N*-acetylcysteamine 177



177

Acid **176** (1 e.g., 215 mg, 1.3 mmol) was dissolved in dry CH_2Cl_2 (18 mL) and cooled to 0 °C. EDCI (1.4 e.q., 295 mg, 1.9 mmol) and DMAP (0.1 e.q., 15 mg, 0.13 mmol) were added. After stirring for 15 minutes at 0 °C, HSNAC **151** (1.4 e.q., 226 mg, 1.9 mmol), in dry CH_2Cl_2 (2 mL), was added and allowed to warm to room temperature and stirred overnight. The reaction was quenched with water (5 mL), the aqueous layer was extracted with CH_2Cl_2 (3 x 20 mL). The organic layers were combined, dried ($MgSO_4$), filtered and concentrated *in vacuo* to give a colourless oil. Purification by flash chromatography (SiO_2, 10% EtOAc/PE, visualised with $KMnO_4$) gave the title compound **177** (222 mg, 0.59 mmol, 46%) as a colourless oil. δ_H (400 MHz, $CDCl_3$); -0.05 (3H, s, $SiCH_3$), -0.03 (3H, s, $SiCH_3$), 0.87 (15H, m, $SiC(CH_3)_3$, , 4-CH_3 & 6-CH_3), 1.08 (3H, d, *J* = 7.1 Hz, 2-CH_3), 1.17-1.42 (5H, m, 5-CH_2, 4-CH), 2.09 (NCH_2), 3.09 (SCH_2), 3.42 (3H, q, *J* = 6.3 Hz, 2-CH), 3.92 (1H, dd, *J* = 7.9, 2.1 Hz, 3-CH) 5.57 (1H, br s, NH); δc (400 MHz, $CDCl_3$); -4.05 ($SiCH_3$), -4.04 ($SiCH_3$), 12.5 (6-CH_3), 13.2 (5-CH_3), 15.0 (4-CH_3), 15.8 (2-CH_3), 18.6 ($COCH_3$), 23.4 (Si<u>C</u>(CH_3), 27.3 (SiC(<u>C</u>H_3), 26.9 (NCH_2), 28.6 (SCH_2), 30.1 (4-CH), 30.2 (2-CH), 170.1 (NCO), 203.5 (SCO); ν_{max} (neat) / cm^{-1}; 3285 (NH), 2959 (Ch), 2857 (CH). 2929 (CH), 1689 (NCO), 1655 (CO), 1251 (CN), 1218 (CN), 1131 (SO), 770 (SiC), 559 (SCO); *m/z* (ESI): 398.21 [$M+Na^+$]; Found (ESI) 398.2159 ($C_{18}H_{37}NNaO_3SSi$ requires 398.2156).

Reference

1. S. E. O'Connor, in *Annual Review of Genetics, Vol 49*, ed. B. L. Bassler, Annual Reviews, Palo Alto, 2015, vol. 49, pp. 71-94.
2. B. Engels, P. Dahm and S. Jennewein, *Metab. Eng.*, 2008, 10, 201-206.
3. A. H. Thomson, C. E. Staatz, C. M. Tobin, M. Gall and A. M. Lovering, *J. Antimicrob. Chemother.*, 2009, 63, 1050-1057.
4. M. Gates and G. Tschudi, *J. Am. Chem. Soc.*, 1956, 78, 1380-1393.
5. R. J. Cox, *Org. Biomol. Chem.*, 2007, 5, 2010-2026.
6. A. Keatinge-Clay, *J. Mol. Biol.*, 2008, 384, 941-953.
7. Y. M. t. Welscher, H. H. t. Napel, M. M. Balagué, C. M. Souza, H. Riezman, B. de Kruijff and E. Breukink, *J. Biol. Chem.*, 2008, 283, 6393-6401.
8. B. K. Law, *Crit. Rev. Oncol. Hematol.*, 2005, 56, 47-60.
9. K. J. Weissman, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*, ed. D. A. Hopwood, 2009, vol. 459, pp. 3-16.
10. C. D. Campbell and J. C. Vederas, *Biopolymers*, 2010, 93, 755-763.
11. I. Fujii, *Nat. Prod. Rep.*, 2009, 26, 155-169.
12. C. R. Valenzano, Y. O. You, A. Garg, A. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2010, 132, 14697-14699.
13. D. Gay, Y. O. You, A. Keatinge-Clay and D. E. Cane, *Biochemistry*, 2013, 52, 8916-8928.
14. Y. Yin, H. Lu, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2003, 125, 5671-5676.
15. K. C. Nicolaou, T. K. Chakraborty, A. D. Piscopio, N. Minowa and P. Bertinato, *J. Am. Chem. Soc.*, 1993, 115, 4419-4420.
16. J. B. Wang, H. X. Pan and G. L. Tang, *Bioorg. Med. Chem. Lett.*, 2011, 21, 3320-3323.
17. D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, 2010, 18, 94-105.
18. J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, 18, 380-416.
19. K. Kinoshita, *J. Synth. Org. Chem. Jpn.*, 2004, 62, 1095-1101.
20. T. Kaneda, *Microbiol. Rev.*, 1991, 55, 288-302.
21. J. E. McMurry and T. P. Bengley, *The Organic Chemistry of Biological Pathways*, Roberts & Co, Englewood, Colorado,, 2005.
22. B. S. Evans, S. J. Robinson and N. L. Kelleher, *Fungal Genet. Biol.*, 2011, 48, 49-61.
23. C. R. Hutchinson, L. Shu-Wen, A. G. McInnes and J. A. Walter, *Tetrahedron*, 1983, 39, 3507-3513.
24. B. Sedgwick, C. Morris and S. J. French, *J. Chem. Soc., Chem. Commun.*, 1978, 193-194.
25. A. Kawaguchi, T. Yoshimura, K. Saito, Y. Seyama, T. Kasama, T. Yamakawa and S. Okuda, *J. Biochem. (Tokyo).* 1980, 88, 1-7.
26. K. Saito, A. Kawaguchi, Y. Seyama, T. Yamakawa and S. Okuda, *Eur. J. Biochem.*, 1981, 116, 581-586.
27. C. Frossl and W. Boland, *J. Chem. Soc., Chem. Commun.*, 1991, 1731-1733.
28. L. C. Du and L. L. Lou, *Nat. Prod. Rep.*, 2010, 27, 255-278.
29. I. B. Lomakin, Y. Xiong and T. A. Steitz, *Cell*, 2007, 129, 319-332.
30. J. Beld, D. J. Lee and M. D. Burkart, *Mol. BioSyst.*, 2015, 11, 38-59.
31. K. Finzel, D. J. Lee and M. D. Burkart, *ChemBioChem*, 2015, 16, 528-547.
32. Y.-H. Chooi and Y. Tang, *J. Org. Chem.*, 2012, 77, 9933-9953.
33. S. Jenni, M. Leibundgut, T. Maier and N. Ban, *Science*, 2006, 311, 1263-1267.
34. T. Maier, S. Jenni and N. Ban, *Science*, 2006, 311, 1258-1262.
35. M. Leibundgut, T. Maier, S. Jenni and N. Ban, *Curr. Opin. Struct. Biol.*, 2008, 18, 714-725.
36. S.-C. Tsai and B. D. Ames, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*, ed. D. A. Hopwood, 2009, vol. 459, pp. 17-47.
37. B. S. Moore and C. Hertweck, *Nat. Prod. Rep.*, 2002, 19, 70-99.

38. H. Ikeda, T. Nonomiya, M. Usami, T. Ohta and S. Ōmura, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, 96, 9509-9514.
39. K. Kakinuma, N. Ikekawa, A. Nakagawa and S. Omura, *J. Am. Chem. Soc.*, 1979, 101, 3402-3404.
40. J. Staunton and B. Wilkinson, *Chem. Rev.*, 1997, 97, 2611-2630.
41. J. F. Martin, *Annu. Rev. Microbiol.*, 1977, 31, 13-38.
42. Y. A. Chan, A. M. Podevels, B. M. Kevany and M. G. Thomas, *Nat. Prod. Rep.*, 2009, 26, 90-114.
43. B. J. Dunn, K. R. Watts, T. Robbins, D. E. Cane and C. Khosla, *Biochemistry*, 2014, 53, 3796-3806.
44. R. J. Cox and T. J. Simpson, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*, ed. D. A. Hopwood, 2009, vol. 459, pp. 49-78.
45. K. M. Fisch, *RSC Advances*, 2013, 3, 18228-18247.
46. K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2007, 8, 289-297.
47. S. Okamoto, T. Taguchi, K. Ochi and K. Ichinose, *Chem. Biol.*, 2009, 16, 226-236.
48. T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, 321, 1315-1322.
49. B. D. Ames, N. Chi, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W. H. Li, J. C. Vederas, Y. Tang and S.-C. Tsai, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, 109, 11144-11149.
50. W. Oppolzer, C. Starkemann, I. Rodriguez and G. Bernardinelli, *Tetrahedron Lett.*, 1991, 32, 61-64.
51. H. Pan, S. C. Tsai, E. S. Meadows, L. J. W. Miercke, A. T. Keatinge-Clay, J. O'Connell, C. Khosla and R. M. Stroud, *Structure*, 2002, 10, 1559-1568.
52. S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Hakansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, 510, 512-517.
53. D. H. Sherman and J. L. Smith, *ACS Chem. Biol.*, 2006, 1, 505-509.
54. A. T. Keatinge-Clay and R. M. Stroud, *Structure*, 2006, 14, 737-748.
55. C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr and D. E. Cane, *Annu. Rev. Biochem.*, 2007, 76, 195-221.
56. J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher and C. A. Townsend, *Science*, 2008, 320, 243-246.
57. J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udwary and C. A. Townsend, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, 103, 16728-16733.
58. J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. A. Hill, O. Kamari-Bidkorpeh, S.-C. Tsai and C. A. Townsend, *Nature*, 2009, 461, 1139-1143.
59. C. R. Huitt-Roehl, E. A. Hill, M. M. Adams, A. L. Vagstad, J. W. Li and C. A. Townsend, *ACS Chem. Biol.*, 2015, 10, 1443-1449.
60. O. Vergnolle, F. Hahn, A. Baerga-Ortiz, P. F. Leadlay and J. N. Andexer, *ChemBioChem*, 2011, 12, 1011-1014.
61. A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, 29, 1050-1073.
62. N. Kandziora, J. N. Andexer, S. J. Moss, B. Wilkinson, P. F. Leadlay and F. Hahn, *Chemical Science*, 2014, 5, 3563-3567.
63. M. Leesong, B. S. Henderson, J. R. Gillig, J. M. Schwab and J. L. Smith, *Structure*, 1996, 4, 253-264.
64. J. Q. Wu, T. J. Zaleski, C. Valenzano, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2005, 127, 17393-17404.
65. J. Wu, T. J. Zaleski, C. Valenzano, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2005, 127, 17393-17404.
66. L. Tang, S. Ward, L. Chung, J. R. Carney, Y. Li, R. Reid and L. Katz, *J. Am. Chem. Soc.*, 2004, 126, 46-47.
67. Y. Li, G. J. Dodge, W. D. Fiers, R. A. Fecik, J. L. Smith and C. C. Aldrich, *J. Am. Chem. Soc.*, 2015, 137, 7003-7006.
68. J. R. Mohrig, K. A. Moerke, D. L. Cloutier, B. D. Lane, E. C. Person and T. B. Onasch, *Science*, 1995, 269, 527-529.

69. C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1998, 120, 2478-2479.
70. L. Moynié, S. M. Leckie, S. A. McMahon, F. G. Duthie, A. Koehnke, J. W. Taylor, M. S. Alphey, R. Brenk, A. D. Smith and J. H. Naismith, *J. Mol. Biol.*, 2013, 425, 365-377.
71. M. S. Kimber, F. Martin, Y. J. Lu, S. Houston, M. Vedadi, A. Dharamsi, K. M. Fiebig, M. Schmid and C. O. Rock, *J. Biol. Chem.*, 2004, 279, 52593-52602.
72. M. Li, Z. Chen, X. Lin, X. Zhang, Y. Song, Y. Wen and J. Li, *Bioorg. Med. Chem. Lett.*, 2008, 18, 5359-5363.
73. K. Stutzman-Engwall, S. Conlon, R. Fedechko, F. Kaczmarek, H. McArthur, A. Krebber, Y. Chen, J. Minshull, S. A. Raillard and C. Gustafsson, *Biotechnol. Bioeng.*, 2003, 82, 359-369.
74. S. Jenni, M. Leibundgut, D. Boehringer, C. Frick, B. Mikolásek and N. Ban, *Science*, 2007, 316, 254-261.
75. D. A. Herbst, R. P. Jakob, F. Zähringer and T. Maier, *Nature*, 2016, 531, 533-537.
76. H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley and R. Lopez, *Nucleic Acids Res.*, 2013, 41, W597-W600.
77. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson and D. G. Higgins, *Mol. Syst. Biol.*, 2011, 7.
78. M. Goujon, H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern and R. Lopez, *Nucleic Acids Res.*, 2010, 38, W695-W699.
79. J. W. Labonte and C. A. Townsend, *Chem. Rev.*, 2013, 113, 2182-2204.
80. T. Schupp, C. Toupet, N. Engel and S. Goff, *FEMS Microbiol. Lett.*, 1998, 159, 201-267.
81. H. G. Floss and T.-W. Yu, *Chem. Rev.*, 2005, 105, 621-632.
82. A. Stratmann, C. Toupet, W. Schilling, R. Traber, L. Oberer and T. Schupp, *Microbiology*, 1999, 145, 3365-3375.
83. T.-W. Yu, Y. Shen, Y. Doi-Katayama, L. Tang, C. Park, B. S. Moore, C. Richard Hutchinson and H. G. Floss, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, 96, 9051-9056.
84. H. G. Floss and T.-W. Yu, *Curr. Opin. Chem. Biol.*, 1999, 3, 592-597.
85. Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman and W. H. Gerwick, *J. Nat. Prod.*, 2004, 67, 1356-1367.
86. X. Guo, T. G. Liu, C. R. Valenzano, Z. X. Deng and D. E. Cane, *J. Am. Chem. Soc.*, 2010, 132, 14694-14696.
87. J D Bergstrom, C Dufresne, G F Bills, a. M Nallin-Omstead and K. Byrne, *Annu. Rev. Microbiol.*, 1995, 49, 607-639.
88. J. D. Bergstrom, M. M. Kurtz, D. J. Rew, A. M. Amend, J. D. Karkas, R. G. Bostedor, V. S. Bansal, C. Dufresne, F. L. VanMiddlesworth, O. D. Hensens and et al., *Proc. Natl. Acad. Sci. U. S. A.*, 1993, 90, 80-84.
89. A. Baxter, B. J. Fitzgerald, J. L. Hutson, A. D. McCarthy, J. M. Motteram, B. C. Ross, M. Sapra, M. A. Snowden, N. S. Watson, R. J. Williams and C. Wright, *J. Biol. Chem.*, 1992, 267, 11705-11708.
90. C. A. Jones, P. J. Sidebottom, R. J. Cannell, D. Noble and B. A. Rudd, *J. Antibiot. (Tokyo).* 1992, 45, 1492-1498.
91. R. J. Cox, F. Glod, D. Hurley, C. M. Lazarus, T. P. Nicholson, B. A. M. Rudd, T. J. Simpson, B. Wilkinson and Y. Zhang, *Chem. Commun.*, 2004, 2260-2261.
92. B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *Chem Commun (Camb)*, 2016, 52, 6777-6780.
93. E. J. Skellam, D. Hurley, J. Davison, C. M. Lazarus, T. J. Simpson and R. J. Cox, *Mol. BioSyst.*, 2010, 6, 680-682.
94. D. M. Roberts, C. Bartel, A. Scott, D. Ivison, T. J. Simpson and R. J. Cox, *Chemical Science*, 2017, 8, 1116-1126.
95. D. Ivison, PhD Thesis, University of Bristol, 2013.
96. D. Roberts, PhD Thesis, University of Bristol, 2014.
97. M. Leibundgut, S. Jenni, C. Frick and N. Ban, *Science*, 2007, 316, 288-290.
98. S. Jackowski and C. O. Rock, *J. Biol. Chem.*, 1987, 262, 7927-7931.

99. Shawn K. Piasecki, Clint A. Taylor, Joshua F. Detelich, J. Liu, J. Zheng, A. Komsoukaniants, Dionicio R. Siegel and Adrian T. Keatinge-Clay, *Chem. Biol.*, 2011, 18, 1331-1340.
100. G. Fráter, U. Müller and W. Günther, *Tetrahedron*, 1984, 40, 1269-1277.
101. Y. Gnas and F. Glorius, *Synthesis*, 2006, 2006, 1899-1930.
102. R. C. Harris, A. L. Cutter, K. J. Weissman, U. Hanefeld, M. C. Timoney and J. Staunton, *J. Chem. Res., Synop.*, 1998, 283.
103. A. D. J., P. I. and S. D. R, *Aldrichim. Acta*, 1997, 30, 3-12.
104. Y. Li, W. D. Fiers, S. M. Bernard, J. L. Smith, C. C. Aldrich and R. A. Fecik, *ACS Chem. Biol.*, 2014, 9, 2914-2922.
105. D. A. Evans, J. Bartroli and T. L. Shih, *J. Am. Chem. Soc.*, 1981, 103, 2127-2129.
106. J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, Oxford University Press, 2nd Edition edn., 2012.
107. K. Hayashi, Y. Hamada and T. Shioiri, *Tetrahedron Lett.*, 1991, 32, 7287-7290.
108. H. Danda, M. M. Hansen and C. H. Heathcock, *J. Org. Chem.*, 1990, 55, 173-181.
109. S. H. Park, S. H. Lee and S. Y. Lee, *J. Chem. Res., Synop.*, 2001, 498-499.
110. T. Kikukawa, M. Imaida and A. Tai, *Bull. Chem. Soc. Jpn.*, 1984, 57, 1954-1960.
111. J. C. Conway, P. Quayle, A. C. Regan and C. J. Urch, *Tetrahedron*, 2005, 61, 11910-11923.
112. F. Sang, D. Li, X. Sun, X. Cao, L. Wang, J. Sun, B. Sun, L. Wu, G. Yang, X. Chu, J. Wang, C. Dong, Y. Geng, H. Jiang, H. Long, S. Chen, G. Wang, S. Zhang, Q. Zhang and Y. Chen, *J. Am. Chem. Soc.*, 2014, 136, 15787-15791.
113. B. P. McKillican, J. W. Perine, D. P. Simmons and T. E. Wilson, *J. Labelled Compd. Radiopharmaceut.*, 2005, 48, 25-30.
114. W. Oppolzer, J. Blagg, I. Rodriguez and E. Walther, *J. Am. Chem. Soc.*, 1990, 112, 2767-2772.
115. W. Oppolzer, C. Starkemann, I. Rodriguez and G. Bernardinelli, *Tetrahedron Lett.*, 1991, 32, 61-64.
116. Thistlethwaite I. R. G, Ph.D, University of Bristol, 2015.
117. R. W. Hoffmann and U. Weidmann, *Chem. Ber.*, 1985, 118, 3966-3979.
118. M. Jacolot, M. Jean, N. Levoin and P. van de Weghe, *Org. Lett.*, 2011, 14, 58-61.
119. J. R. Gage and D. A. Evans, *Org. Synth.*, 1990, 68, 83-91.
120. V. P. Kumar and S. Chandrasekhar, *Org. Lett.*, 2013, 15, 3610-3613.
121. N. Schlager and A. Kirschning, *Org. Biomol. Chem.*, 2012, 10, 7721-7729.
122. C. Neri and J. M. J. Williams, *Adv. Synth. Catal.*, 2003, 345, 835-848.
123. L. Pouységu, M. Marguerit, J. Gagnepain, G. Lyvinec, A. J. Eatherton and S. Quideau, *Org. Lett.*, 2008, 10, 5211-5214.
124. K. Terpe, *Appl. Microbiol. Biotechnol.*, 2006, 72, 211-222.
125. D. M. Francis and R. Page, in *Current Protocols in Protein Science*, John Wiley & Sons, Inc., 2010.
126. J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning a Laboratory Manual*, SECOND EDITION edn., 1989.
127. C. P. Papaneophytou and G. Kontopidis, *Protein Expression Purif.*, 2014, 94, 22-32.
128. I. Lee and C. K. Suzuki, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2008, 1784, 727-735.
129. T. A. Phillips, R. A. VanBogelen and F. C. Neidhardt, *J. Bacteriol.*, 1984, 159, 283-287.
130. J. Haiko, M. Suomalainen, T. Ojala, K. Lahteenmaki and T. K. Korhonen, *Innate Immunity*, 2009, 15, 67-80.
131. L. Dumon-Seignovert, G. Cariot and L. Vuillard, *Protein Expression Purif.*, 2004, 37, 203-206.
132. S. B. Mulrooney and L. Waskell, *Protein Expression Purif.*, 2000, 19, 173-178.
133. S. E. Bondos and A. Bicknell, *Anal. Biochem.*, 2003, 316, 223-231.
134. A. P. Golovanov, G. M. Hautbergue, S. A. Wilson and L.-Y. Lian, *J. Am. Chem. Soc.*, 2004, 126, 8933-8939.
135. J. A. Vasina and F. Baneyx, *Protein Expression Purif.*, 1997, 9, 211-218.
136. S. Sahdev, S. K. Khattar and K. S. Saini, *Mol. Cell. Biochem.*, 2008, 307, 249-264.

137. C. P. Chou, *Appl. Microbiol. Biotechnol.*, 2007, 76, 521-532.
138. C. P. Papaneophytou, V. Rinotas, E. Douni and G. Kontopidis, *Protein Expression Purif.*, 2013, 90, 9-19.
139. T. Schagat, *Promega Notes*, 2007, 96, 22-23.
140. P. N. Hengen, *Trends Biochem. Sci.*, 1995, 20, 285-286.
141. J. Jancarik, Pufan Ramona, C. Hong, S.-H. Kima and R. Kima, *Acta Crystallogr.*, 2007, D60, 1670-1673.
142. C. C. Aldrich, B. J. Beck, R. A. Fecik and D. H. Sherman, *J. Am. Chem. Soc.*, 2005, 127, 8441-8452.
143. K. A. Johnson and R. S. Goody, *Biochemistry*, 2011, 50, 8264-8269.
144. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, 25, 3389-3402.
145. D. A. Herbst, R. P. Jakob, F. Zähringer and T. Maier, *Nature*, 2016, 531, 533.
146. A. Faille, Slama, N., Quemard, A., Mourey, L., Pedelacq, J.D., *PDB ID: 400C; Insights into the catalytic mechanism of the DH domain of the Mycobacterium tuberculosis polyketide synthase PpsC and architecture of the beta-carbon processing domains*, http://www.rcsb.org/pdb/explore/explore.do?structureId=4OOC.
147. W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso and R. Lopez, *Nucleic Acids Res.*, 2015, 43, W580-W584.
148. A. Roy, A. Kucukural and Y. Zhang, *Nat. Protocols*, 2010, 5, 725-738.
149. M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. M. and and A. Šali, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, 29, 291-325.
150. J. Peng and J. Xu, *Proteins: Struct., Funct., Bioinf.*, 2011, 79, 161-171.
151. A. Fiser and A. Šali, in *Methods Enzymol.*, eds. Charles W. Carter, Jr. and M. S. Robert, Academic Press, 2003, vol. Volume 374, pp. 461-491.
152. D. E. Kim, D. Chivian and D. Baker, *Nucleic Acids Res.*, 2004, 32, W526-W531.
153. M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede, *Nucleic Acids Res.*, 2014, 42, W252-W258.
154. J. Yang and Y. Zhang, *Curr Protoc Bioinformatics*, 2015, 52, 5.8.1-15.
155. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Protocols*, 2015, 10, 845-858.
156. M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu and J. Xu, *Nat. Protocols*, 2012, 7, 1511-1522.
157. S. Raman, R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. DiMaio, O. Lange, L. Kinch, W. Sheffler, B.-H. Kim, R. Das, N. V. Grishin and D. Baker, *Proteins: Struct., Funct., Bioinf.*, 2009, 77, 89-99.
158. D. E. Kim, D. Chivian and D. Baker, *Nucleic Acids Res.*, 2004, 32, W526-W531.
159. Y. Song, F. DiMaio, Ray Y.-R. Wang, D. Kim, C. Miles, T. J. Brunette, J. Thompson and D. Baker, *Structure*, 2013, 21, 1735-1742.
160. J. Söding, A. Biegert and A. N. Lupas, *Nucleic Acids Res.*, 2005, 33, W244-W248.
161. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas and V. Alva, *J. Mol. Biol.*, DOI: https://doi.org/10.1016/j.jmb.2017.12.007.
162. P. Benkert, M. Biasini and T. Schwede, *Bioinformatics*, 2011, 27, 343-350.
163. P. Benkert, T. Schwede and S. C. E. Tosatto, *BMC Struct. Biol.*, 2009, 9, 35-35.
164. P. Benkert, S. C. E. Tosatto and D. Schomburg, *Proteins: Struct., Funct., Bioinf.*, 2008, 71, 261-277.
165. G. Studer, M. Biasini and T. Schwede, *Bioinformatics*, 2014, 30, i505-i511.
166. D. A. Evans, *Angewandte Chemie International Edition*, 2014, 53, 11140-11145.
167. E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus, *Proteins*, 2009, 77, 114-122.
168. S. Smith, A. Witkowski and A. K. Joshi, *Prog. Lipid Res.*, 2003, 42, 289-317.
169. A. B. Pangborn, M. A. Giardello, R. H. Grubbs, R. K. Rosen and F. J. Timmers, *Organometallics*, 1996, 15, 1518-1520.
170. P. Renaud and D. Seebach, *Helv. Chim. Acta*, 1986, 69, 1704-1710.

171. W. Yuan, Y. Jia, J. Tian, K. D. Snell, U. Müh, A. J. Sinskey, R. H. Lambalot, C. T. Walsh and J. Stubbe, *Arch. Biochem. Biophys.*, 2001, 394, 87-98.
172. M. S. Lee, G. W. Qin, K. Nakanishi and M. G. Zagorski, *J. Am. Chem. Soc.*, 1989, 111, 6234-6241.
173. S. Klopries, U. Sundermann and F. Schulz, *Beilstein J. Org. Chem.*, 2013, 9, 664-674.
174. S. Chang, S. Hur and R. Britton, *Angewandte Chemie International Edition*, 2015, 54, 211-214.
175. A. E. May, P. H. Willoughby and T. R. Hoye, *J. Org. Chem.*, 2008, 73, 3292-3294.
176. J. Peed, I. Periñán Domínguez, I. R. Davies, M. Cheeseman, J. E. Taylor, G. Kociok-Köhn and S. D. Bull, *Org. Lett.*, 2011, 13, 3592-3595.
177. F. Ding, M. L. Leow, J. Ma, R. William, H. Liao and X.-W. Liu, *Chemistry – An Asian Journal*, 2014, 9, 2548-2554.
178. A. D. Fotiadou and A. L. Zografos, *Org. Lett.*, 2011, 13, 4592-4595.
179. U. Ramulu, D. Ramesh, S. Rajaram, S. P. Reddy, K. Venkatesham and Y. Venkateswarlu, *Tetrahedron: Asymmetry*, 2012, 23, 117-123.
180. D. Brandt, A. Dittoo, V. Bellosta and J. Cossy, *Org. Lett.*, 2015, 17, 816-819.
181. G. V. M. Sharma and P. S. Reddy, *Eur. J. Org. Chem.*, 2012, 2012, 2414-2421.
182. H. Ren and W. D. Wulff, *Org. Lett.*, 2012, 15, 242-245.
183. Q. Yue, Y. Zhao, B. Sun, L. Hai, L. Guo and Y. Wu, *Chin. J. Chem.*, 2015, 33, 1145-1152.
184. K. Sugimoto, Y. Kobayashi, A. Hori, T. Kondo, N. Toyooka, H. Nemoto and Y. Matsuya, *Tetrahedron*, 2011, 67, 7681-7685.
185. L. Liu, R. S. Tanke and M. J. Miller, *The Journal of Organic Chemistry*, 1986, 51, 5332-5337.
186. R. Moumné, B. Denise, K. Guitot, H. Rudler, S. Lavielle and P. Karoyan, *Eur. J. Org. Chem.*, 2007, 2007, 1912-1920.

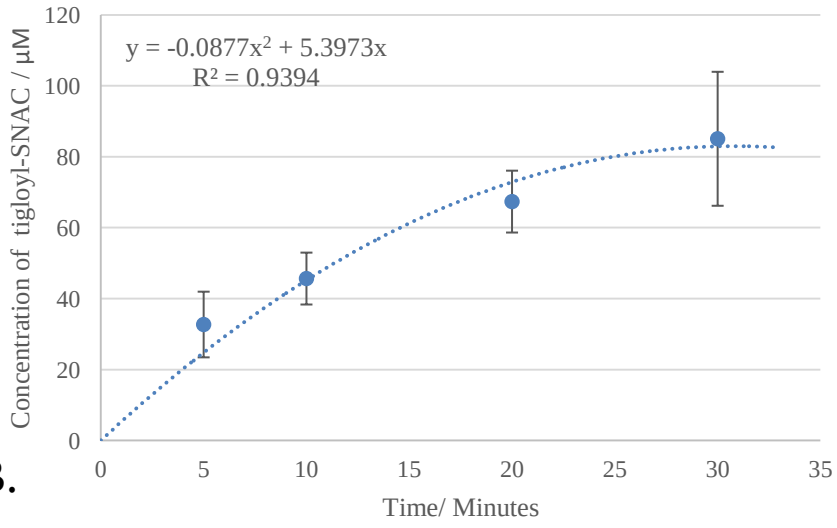
Appendix 1 – CLUSTAL Omega full multiple sequence alignment

```
SQTKS_DH          ------------------------------------------NKAHRQRVHPPHDLLGSL          18
2vz8_mFASDH       PPVEFPAPRGTPLISPHXKWDHSQAWDVPSA-------ADF---PSGS------------          38
4LN9_RifDH10      ------------MGSSHHHHHHSSGL-VPRGSHMADTASDAVSLGLAGA---DHPLLGAV          44
3EL6_EryDH4       -----------EMGSSHHHHHHSSGL-VPRGSHMHRP-ADVSALGVRGA---EHPLLLAA          44
3KG7_CurHDH       ------------------------------------------SNASGQQ---VHRLLGNK          15
3KG8_CurJ_DH      ------------------------------------------SNASTTK---LHPLINQK          15
3KG6_CurF_DH      ----------------------------------------------SN---AHPLLGEK          10
3KG9_CurK_DH      ------------------------------------------SNATAKN---LHPLLGEK          15


SQTKS_DH          IVG-RDLREPTWRHFIRVQD-IPWIRDHVVQSALVYPGAGFICMAMEAMVQLHELRDSQS          76
2vz8_mFASDH       ----SCSSVAVYKFDVSPESPDHYLVDHCIDGRVLFPGTGYLWLTWKTLARALSQNL---          91
4LN9_RifDH10      VQL-PQSDGLVFTSRLSLRS-HPWLADHAVRDVVIVPGTGLVELAVRAGDEA--------          94
3EL6_EryDH4       VDV-PGHGGAVFTGRLSTDE-QPWLAEHVVGGRTLVPGSVLVDLALAAGEDV--------          94
3KG7_CurHDH       LELASTGQ-TIYHQDINLNN-HPWIGDHRVYDTPVIPGVSYIAMTLAAV-----------          62
3KG8_CurJ_DH      FQS-PLSKEIFFESYFSTEN-LPFLADHIVYEQVVVPGASHISLLLAAASLTF-------          66
3KG6_CurF_DH      INLAGIEDQHRFQSYIGAES-PGYLNHHQVFGKVLFPSTGYLEIAASAGKSLF---T---          63
3KG9_CurK_DH      LNLARIENQHHFQSYLTAES-PAYLSQHQVFNKVLFPATGYLEIAAAVGKNLL---T---          68
                            :   .   .   :: .* :    : *.   : :   .

SQTKS_DH          RKVAGYRLAEVDILRAMLIPDTSEGLEAHISLRPCSTKLLLTNEWYDFCVSSVGD-----          131
2vz8_mFASDH       -EETPVVFEDVTLHQATILPKTGTV-SLEVRLLEA----------SHAFEVSDS-------          133
4LN9_RifDH10      ---GCPVLDELVIEAPLVVPRRGGV-RVQVALGGP-----ADDGSRTVDVFSLRED----          141
3EL6_EryDH4       ---GLPVLEELVLQRPLVLAGAG-A-LLRMSVGAP-----DESGRRTIDVHAAEDV----          140
3KG7_CurHDH       --GVPAAVEDINFQQPLFLAESNTTRETQLMLHTA------DNVGKQF------------          102
3KG8_CurJ_DH      -AATECQIEDILFPQALAIPEQGVR-TVQVVLT-P------QNNSFSFQVISFDDSLESQ          117
3KG6_CurF_DH      -SQEQVVVSDVDILQSLVIPETEIK-TVQTVVSFA------ENNSYKFEIFSPSEG----          111
3KG9_CurK_DH      -TGEQVVVSDVTIVRGLVIPETDIK-TVQTVISTL------ENNSYKLEIFSTSEG----          116
                         . :: :     :         .  :               .

SQTKS_DH          ----------DDKFVDHCRGRITIEFDTSGSADTPRTSLRERSRSTGLMRSVDPSNLYSF          181
2vz8_mFASDH       ------------NGSLIASGKVYQWESPDPKLFDTRAAVD--PADSTAEFRLSQGDVYKD          179
4LN9_RifDH10      ----------ADSWLRHATGVLVPENRPRGTAAFDFAAW---PPPEAK--PVDLTGAYDV          186
3EL6_EryDH4       ------ADLADAQWSQHATGTLAQGV---AAGPRDTEQW---PPEDAV--RIPLDDHYDG          186
3KG7_CurHDH       VEVFSRDGAKQEEWQQHASMSVSENPPPPPTLSVD---I---PALCEQLRPLDTDTLTEI          156
3KG8_CurJ_DH      INQVSNNGSHISDWAVHATGKLSVAN--AEQSLIPLEEI---QARCSQ--KIDSAEIYQH          170
3KG6_CurF_DH      ------ENQQTPQWVLHAQGKIYTEPTRNSQAKIDLEKY---QAECSQ--AIEIEEHYRE          160
3KG9_CurK_DH      ------DNQQANQWTLHAEGKIFLDSTTNTKAKIDLEQY---QRECSQ--VIDIQQHYQQ          165
                            .    .   :                          :

SQTKS_DH          LRAQGIYHGPIFQNLKTISSRKDHSE----SSFVVANTASVMPNGFQSPHVIHPTTLDSI          237
2vz8_mFASDH       LRLRGYDYGPFFQLVLESDLEGNRG------RLQ--------------WNDSWVSFLDAM          219
4LN9_RifDH10      LADVGYGYGPTFRAVRAVWRRGSGNTTETFAEIALPEDARAE----AGRFGIHPALLDAA          242
3EL6_EryDH4       LAEQGYEYGPSFQALRAAWRKDDSV----YAEVSIA----AD----EEGYAFHPVLLDAV          234
3KG7_CurHDH       YASISLVYGPMLQAVRQAWIGEETS----LLEIEVPKALAFQ----LAGEPIHPVLIDAC          208
3KG8_CurJ_DH      LWDRQIHLGQSFRWIEQVWLGEGEV----LCQMKVPKTILN-----TTKYQLHPTLVDSC          221
3KG6_CurF_DH      YRSKGIDYGSSFQGIKQLWKGQGKA----LGEMAFPEELTAQ----LADYQLHPALLDAA          212
3KG9_CurK_DH      FKSRGIDYGNSFQGIKQLWKGQGKA----LGKIALPEEIAGQ----ATDYQLHPALLDAA          217
                         *  :: :                .                      :*:

SQTKS_DH          FQGAYTALPGAG-LDQNTAMIPRSIQELYLSSALTSDVGQCLVSDTSLIRYDGQSFTVNV          296
2vz8_mFASDH       LHMSIL-AP-----GQLGLYLPTRFTSIRIDPVTHRQKLYTLQDTTQAAD---VVVDRNL          270
4LN9_RifDH10      LHSTMVSAAADTESYGDEVRLPFAWNGLRLHAAGASV-LRVRVAKP-------ERDSLSL          294
3EL6_EryDH4       AQTLSLGALG----EPGGGKLPFAWNTVTLHASGATS-VRVVATPA-------GADAMAL          282
3KG7_CurHDH       TRLT-PDLFDFS-SDSGVFWAPWRVKEMTLSHPTPSR-FYAYVEEPSRVNEQLQTRSYDI          265
3KG8_CurJ_DH      FQSIIALVLDQS-GNKNETFVPFSIDKFTFYNSSDNDLLWCYTCG-SKDKQSGEKFKADI          279
3KG6_CurF_DH      FQIV-SYAIPHT-ET-DKIYLPVGVEKFKLYRQTISQ-VWAIAE------IRQTNLTANI          262
3KG9_CurK_DH      LQIL-GHAIGNT-ETDDKAYLPVGIDKLKQYRQTITQ-VWAIVE------IPENTLKGSI          268
                   :                  *    .                          :

SQTKS_DH          DVSSKADSEHTPVLEIKGLRNQSVGQMA--------------------------------          324
2vz8_mFASDH       NTVV---AGGALFLGAHSSVAPRRPQEHLKPILEKFCFTPHVESGCLAGNTALQEELQLC          327
4LN9_RifDH10      EAVD---ESGGLVVTLDSLVGRPVSNDQLTTA---------------------------A          324
3EL6_EryDH4       RVTD---PAGHLVATVDSLVVRSTGEKWEQPE---------------------------P          312
3KG7_CurHDH       QLLD---ETGQAFGRINGFTVKRAPSQLFLK-----------------------------          293
3KG8_CurJ_DH      QLFD---QHGQLVAQVIGFEGRKANPKILLMT----------------------------          308
3KG6_CurF_DH      FLVD---NQGTVLVELEGLRVKVTEP----------------------------------          285
3KG9_CurK_DH      KLVD---NQGSLLAEIEGLRVTATTADALLK-----------------------------          296
                         .    .

SQTKS_DH          ----------------------------          324
2vz8_mFASDH       RGLAQALQTKVAQQGLKMVVPGLDGAQ          354
4LN9_RifDH10      G---------------------------          325
3EL6_EryDH4       RG--------------------------          314
3KG7_CurHDH       ----------------------------          293
3KG8_CurJ_DH      ----------------------------          308
3KG6_CurF_DH      ----------------------------          285
3KG9_CurK_DH      ----------------------------          296
```

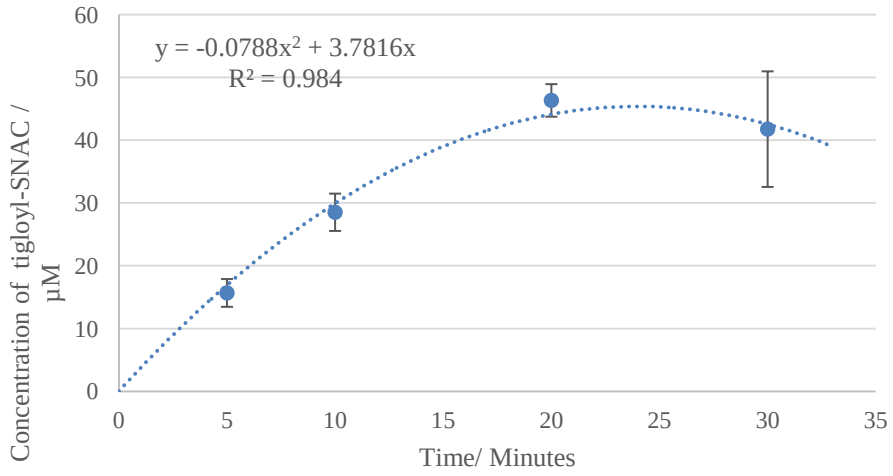
Appendix 2 – Kinetic data for substrate 2*R*,3*R* -114

Graphs produced using Microsoft excel. Automatic second order polynomial trendlines fitted using microsft excel curve fitting tool.

A.



B.



C.

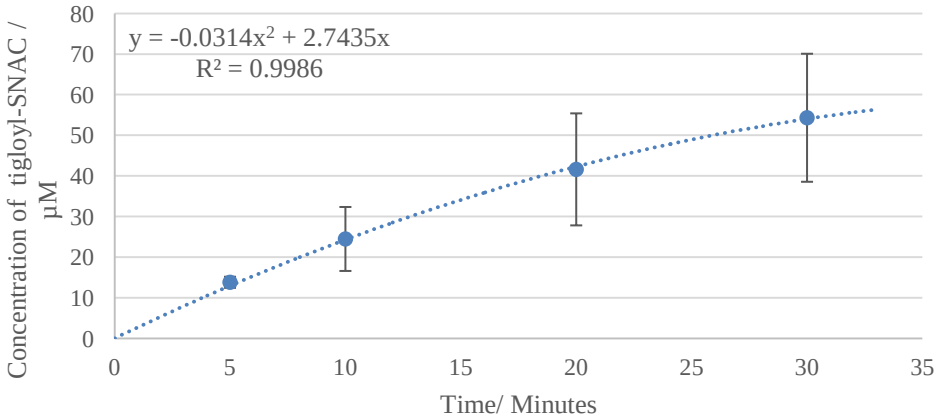


Figure 89 – Initial rate of dehydration of **114** by the isolated DH domain from SQTKS to form 112: A. 114 concentration 4.5 mM; B. 114 concentration 3.0 mM; C. 114 concentration 2.3 mM.

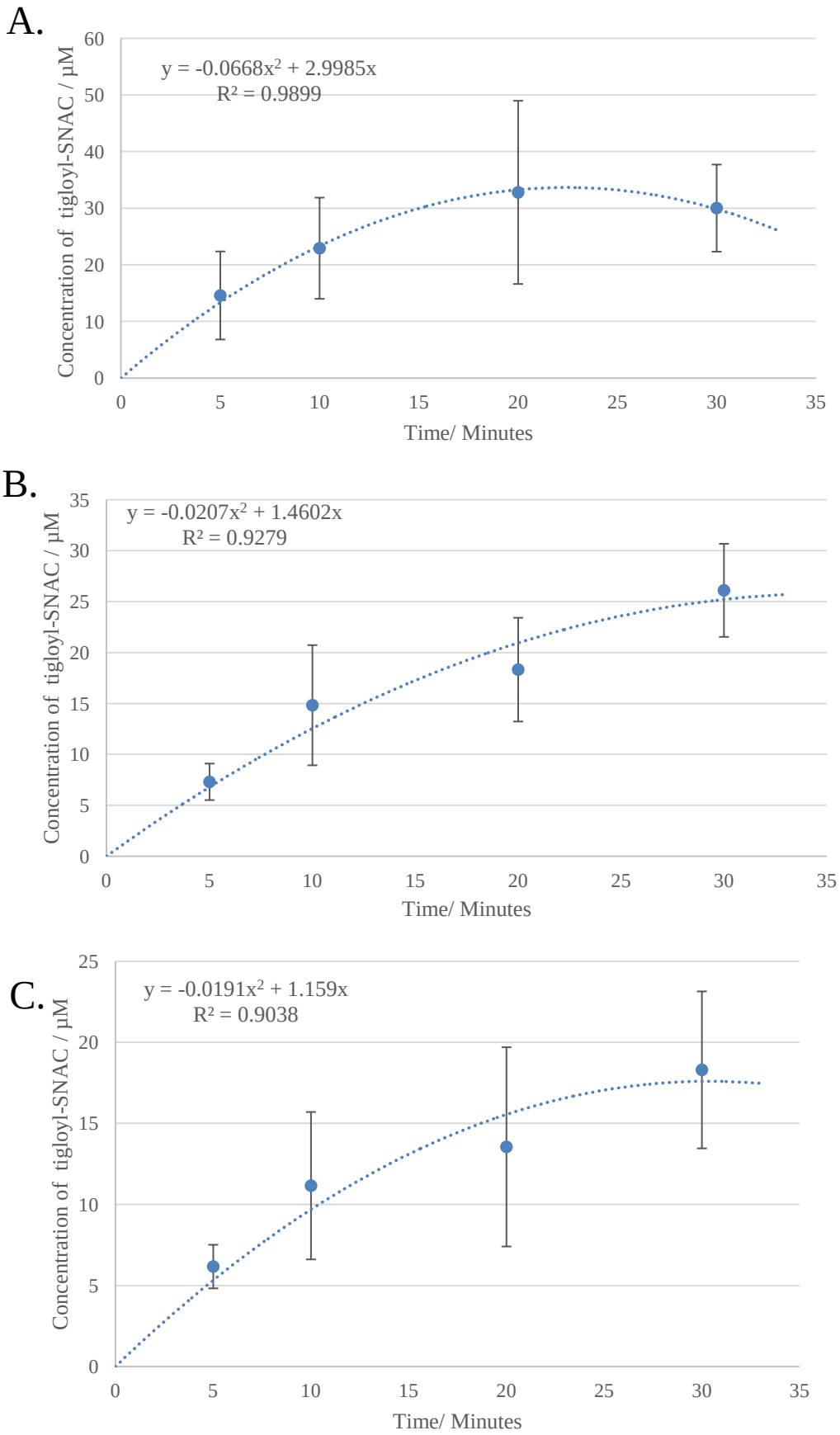


Figure 90 Initial rate of dehydration of **114** by the isolated DH domain from SQTKS to form **112**: **A**. **114** concentration 1.0 mM; **B**. **114** concentration 0.8 mM.; **C**. **114** concentration 0.5mM

A.

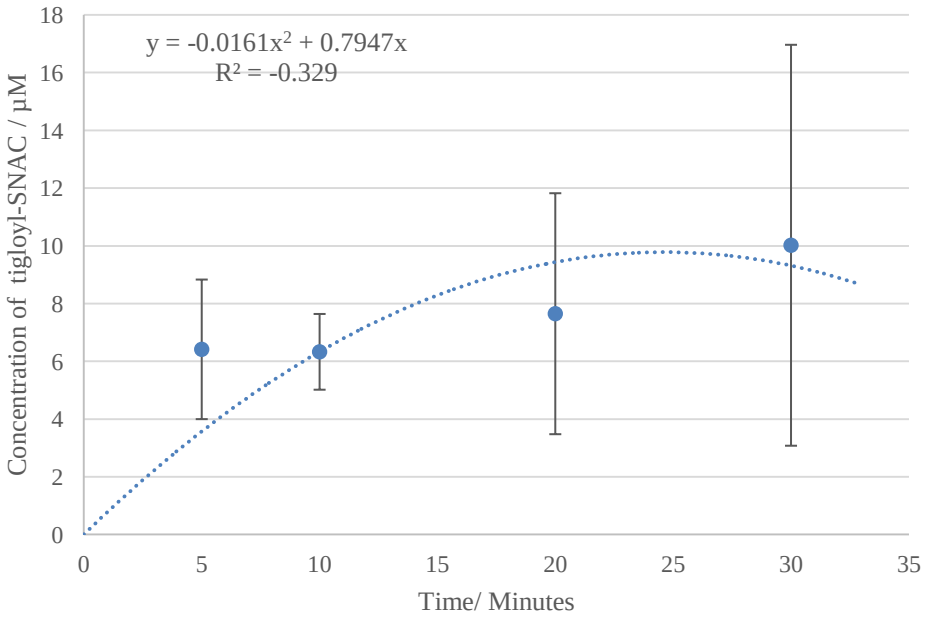
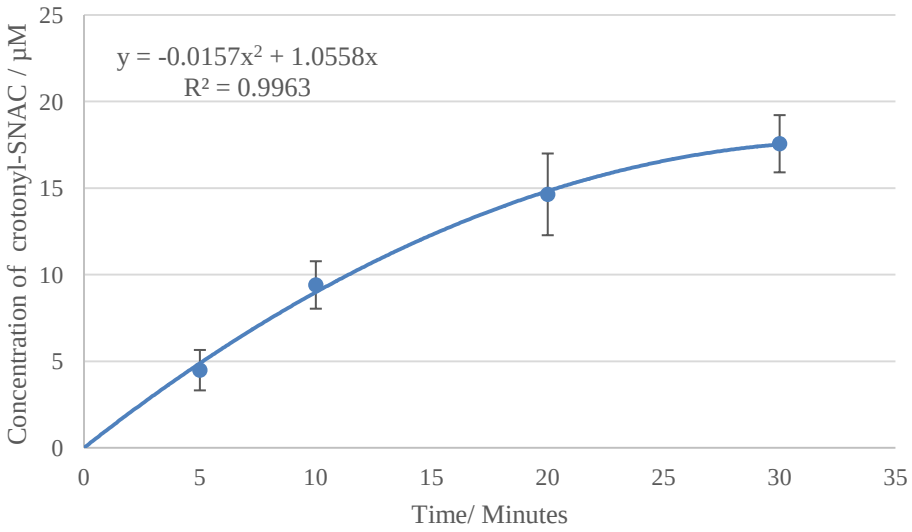


Figure 91 Initial rate of dehydration of **114** by the isolated DH domain from SQTKS to form **112**: **A**. **114** concentration 0.2 mM.

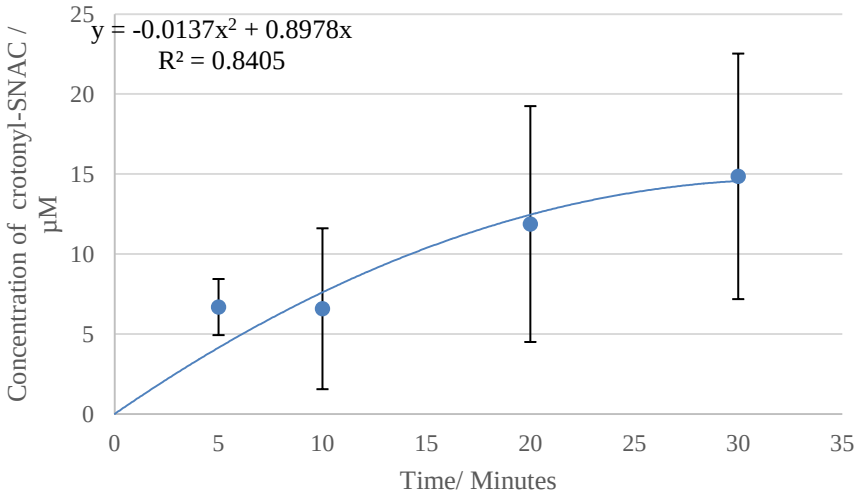
Appendix 3 - Kinetic data for substrate 3*R* -114

Graphs produced using Microsoft excel. Automatic second order polynomial trendlines fitted using microsft excel curve fitting tool.

A.



B.



C.

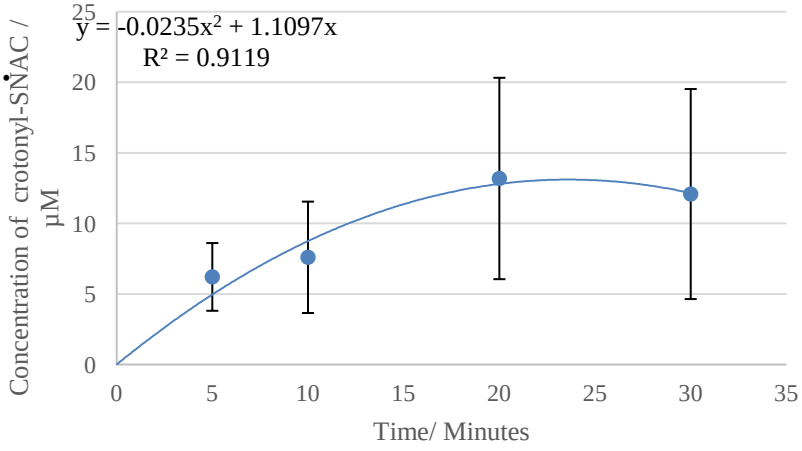
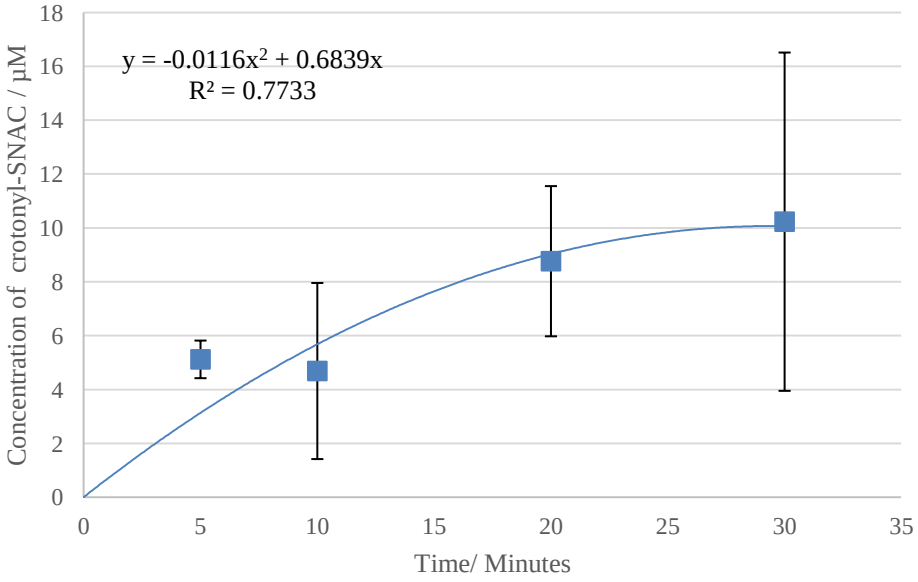
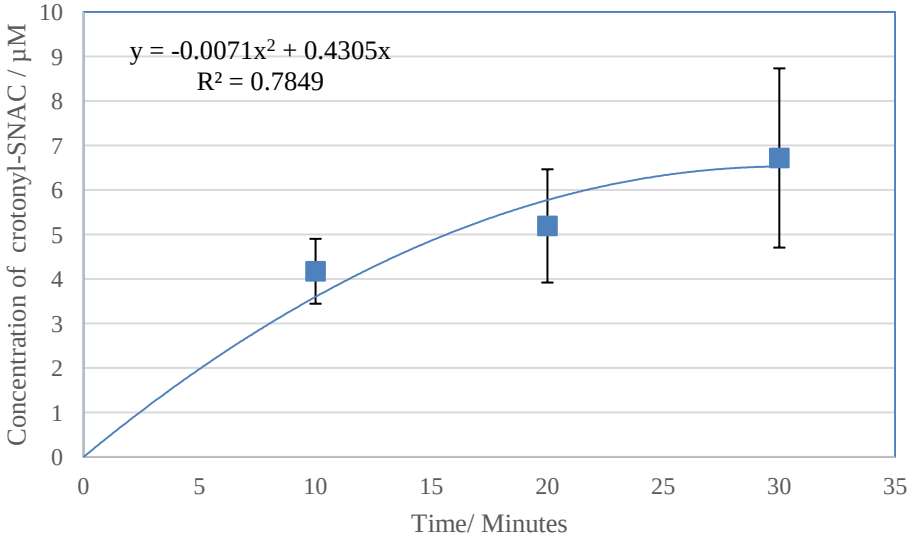


Figure 92 Initial rate of dehydration of **124** by the isolated DH domain from SQTKS to form **185**: **A**. **124** concentration 3.99 mM; **B**. **124** concentration 3.50 mM, **C**. **124** concentration 2.6 mM.

A.



B.



C.

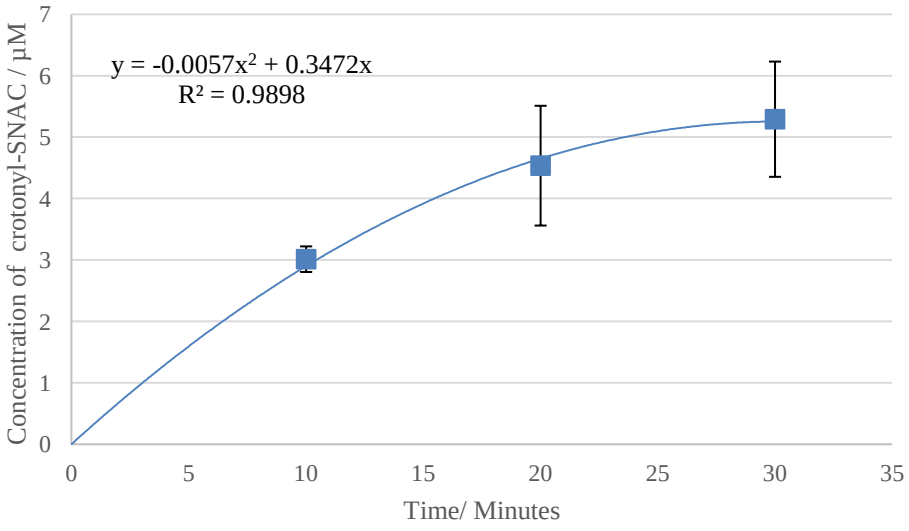


Figure 93 Initial rate of dehydration of **124** by the isolated DH domain from SQTKS to form **185**: **A**. X concentration 2.0 mM; **B**. **124** concentration 0.88 mM, **C**. **124** concentration 0.64 mM.

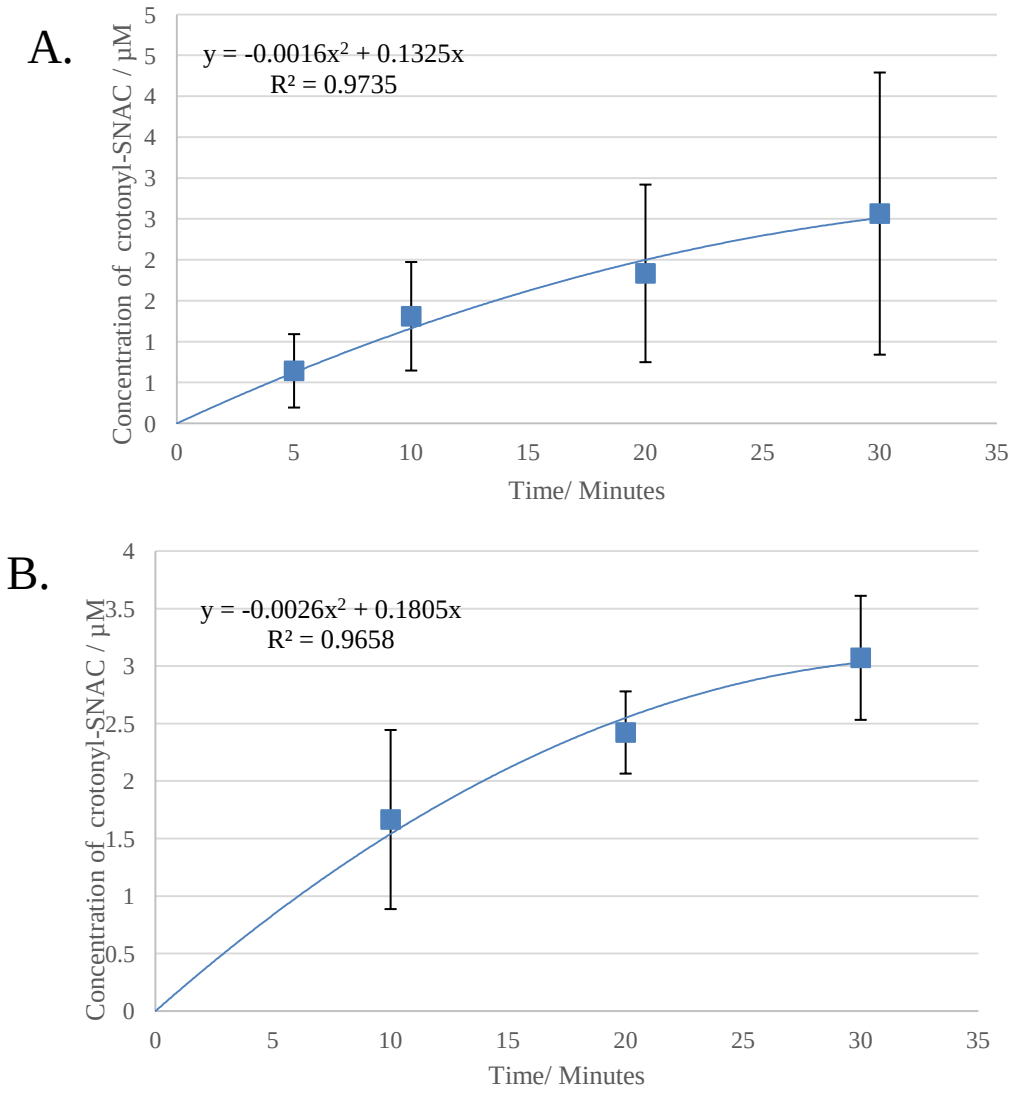


Figure 94 Initial rate of dehydration of **124** by the isolated DH domain from SQTKS to form **185**: **A**. **124** concentration 0.44 mM; **B**. **124** concentration 0.20 mM.